(72) Inventors; and
(75) Inventors/Applicants (for US only): ECKER, David,
J. [US/US]; 1041 Saxony Road, Encinitas, CA 92024
(US). GRIFFEY, Richard, H. [US/US]; 360 Barsby
Street, Vista, CA 92084 (US). SAMPATH, Rangarajan
[IN/US]; 12223 Mannix Road, San Diego, CA 92129 (US).
HOFSTADLER, Steven, A. [US/US]; 5014 Viewridge
Way, Oceanside, CA 92056 (US). MCNEIL, John
[US/US]; 6389 Castejon Drive, La Jolla, CA 92037 (US).
CROOKE, Stanley, T. [US/US]; 3211 Piragua Street,
Carlsbad, CA 92009 (US). BLYN, Lawrence, B. [US/US];
27892 Perales, Mission Viejo, CA 92692 (US). HALL,
Thomas, A. [US/US]; 5239 Wohlford Street, Oceanside,
CA 92056 (US).

(54) Title: METHODS FOR RAPID DETECTION AND IDENTIFICATION OF BIOAGENTS FOR ENVIRONMENTAL AND PRODUCT TESTING
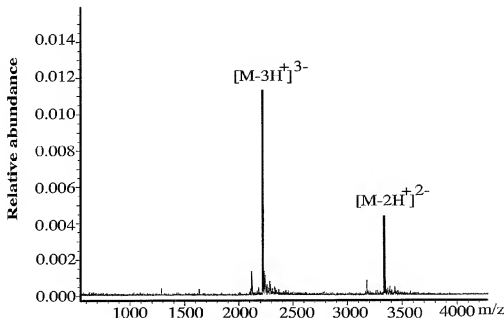
(57) Abstract: The present invention provides methods for rapid detection of bioagents for environmental and product testing. The methods can be used for testing air, water, soil, surfaces of buildings, containers, towers and the like, as well as testing of foodstuff and cosmetics.

WO 2004/053076 A2

MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE,

SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS FOR RAPID DETECTION AND IDENTIFICATION OF BIOAGENTS FOR ENVIRONMENTAL AND PRODUCT TESTING

FIELD OF THE INVENTION

This invention relates to the field of environmental and product testing. The invention provides rapid detection and identification of bioagents in environmental and product samples.

BACKGROUND OF THE INVENTION

Testing for potentially harmful microbes in both natural and human-modified environments represents a significant public health challenge. Microbial contamination of waters can lead to numerous serious issues. Bacterial contamination can cause adverse economic impacts for the affected areas by closing popular recreational areas for extended periods of time. This is particularly relevant in the heavily visited recreational waters and the swimming beaches in urban centers. In addition, bacterial contamination can cause severe illnesses such as gastrointestinal disorders. Microbial contaminants may not be limited to surface or recreational waters. In fact, a far more costly migration of these microbial contaminants is their movement to groundwater supplies which are used for drinking water.

The cysts of *Cryptosporidium* are of increasing importance because of their presence in water supplies. When in the gut, four spindle-shaped motile sporozooites burst from the cyst to infect gut epithelial cells and continue their life cycle. *Entamoeba histolytica*, another water-borne pathogen, can cause diarrhea or a more serious invasive liver abscess. When in contact with human cells, these amebae are cytotoxic. *Giardia* is found in contaminated rivers and lakes and is also be contracted via contaminated foods. There is some evidence that a heavy infection of attached Giardia physically blocks the important transport of nutrients across the epithelium (see, for example, www.cellsalive.com/parasit.htm).

Detection of airborne pathogens has also garnered much attention in light of the recent anthrax attack where anthrax spores were disseminated from mail packages.

A recent news report has stated that the anthrax mailed to a Senate office last fall was able to become airborne again even after it settled in the office. The fact that ordinary movement in the office was enough to send anthrax back into the air provides evidence that the spores were altered to make them more dangerous. Scientists thought that once the anthrax-laden envelope was opened, the spores would settle in the office and would be unlikely to become airborne again. But simulation of normal workplace activities such as paper handling, walking around the office and mail sorting several weeks after the envelope arrived, caused the spores to become

airborne. Measurements of anthrax in the air of the office were substantially higher after researchers simulated everyday office activity than when the office was more still. And more than 80% of the airborne anthrax spores were of a size that could easily be breathed into the lungs (see, for example, www.ph.ucla.edu/epi/bioter/officeactivityanthrax.html).

5   According to a study published in 1999 by the Centers for Disease Control and Prevention (Atlanta), there were almost 14 million cases of foodborne illness caused by known pathogenic microbes (bacteria, parasites, and viruses) in the United States in 1997, and these illnesses caused more than 60,000 hospitalizations and almost 2000 deaths (1). Although ~70% of the reported illnesses were caused by food contaminated with Norwalk-like viruses, almost

10 half of the deaths were the result of infection with Listeria monocytogenes and Salmonella species bacteria. Foodborne microbial contamination comes in a variety of guises and from myriad sources. Perhaps the best-known foodborne illnesses come from the various subspecies of *E. coli*. Largely associated with raw or undercooked ground beef, *E coli*. infection leads to gastrointestinal disorders. Infection is usually self-limited and can last for about eight days. In

15 some cases, however, infection can lead to hemolytic uremic syndrome, which causes renal failure and anemia, and it can be deadly to children, the elderly, and those with weakened immune systems. Certain subtypes of *E. coli* can cause severe, even lethal illness. For example, in the spring of 1992-93, a multi-state outbreak of over 500 cases of *E. coli* O157:H7 infections was associated with a restaurant chain (www.dhss.state.mo.us/MoEpi/moepi161.pdf).

20   One of the least understood but increasingly prevalent illnesses is caused by *Listeria monocytogenes*. Associated with raw milk, soft cheeses, and raw meats, L. monocytogenes infection can lead to meningitis, encephalitis, and intrauterine or cervical infections in pregnant women that can cause spontaneous abortion or stillbirth. *Listeria* is able to grow under a wide variety of conditions, including refrigeration.

25   Another common infection is caused by *Salmonella* species. Associated with raw meats, poultry, and eggs, *Salmonella* infection can lead to nausea, vomiting, cramps, and diarrhea. Typically, acute symptoms subside after a couple of days. In severe cases of *S. typhi* or *paratyphi*, however, infection can cause a typhoid-like fever and possible septicemia.

           As with many other bacteria, *Campylobacter jejuni* infections also lead to

30 gastrointestinal disorders. Infection is usually self-limiting and lasts 7–10 days. Associated largely with poultry, *Campylobacter* is also found in unchlorinated water but can be eliminated by boiling.

           The family of Norwalk-like viruses (NLVs) is spread through feces-contaminated water, and are thus largely associated with shellfish. Although most people have been exposed to NLVs

-3-

at some point in their lives, they rarely exhibit any symptoms. Disease associated with infection is usually mild, with symptoms of vomiting, nausea, and abdominal pain, but infection is usually self-limiting and symptoms subside within 2 days.

Corporations and government health officials have put much effort into food testing to prevent human suffering and avoid lawsuits. The two big challenges to large-scale testing of samples, both during food processing and from a clinical perspective, are time and sensitivity. Traditional microbiological methods require that the microbes be cultured and characterized for a variety of metabolic and physical markers. This process can take days to weeks, depending on the organism. For example, in a test to detect *Salmonella*, one bacterial pre-enrichment step takes 16–20 h, a *Salmonella*-specific enrichment takes another 24 h, and a final identification step in which cultures are streaked out onto selective media can take 24–48 h. If the results are positive, they must then be confirmed by sub-cultivation and serological testing. Thus, this assay can take anywhere from 3 to 6 days. During this time, the food products might decay beyond the selling point. Consumers might contract food poisoning from the products, putting them at risk for sickness or death. It is therefore preferable to have an assay that can locate and identify the offending microbes in a timeframe that is measured in hours, not days.

Similarly, the assay must also be very sensitive because the tested samples might have no more (and possibly less) than one cell per milliliter or gram of starting material. *Listeria* infections can start from as few as 10 cells. As with most traditional microbiological tests, this requires some form of microbe culture enrichment using a growth medium, but it is critical that this step not take too much time (see, for example, pubs.acs.org/subscribe/journals/tcaw/12/i03/pdf/303willis.pdf).

Bacteria and their enzymes, along with some fungi and critical nutrient additives are cost effective agents for in-situ remediation (otherwise known as bioremediation) of hazardous wastes and subsurface pollution in soils, sediments and wastewaters. The ability of each bacterial strain to degrade toxic waste depends on the nature of each contaminant. Since most sites are typically comprised of multiple pollutant types, the most effective approach to bioremediation is to use a mixture of bacterial species/strains, each specific to the degradation of one or more types of contaminants. It is critical to monitor the composition of the indigenous and added bacterial consortium in order to evaluate the activity level of the bacteria, and to permit modifications of the nutrients and other conditions for optimizing the bioremediation process. Additionally, it is desirable to return a bioremediation site to its natural state following the bioremediation process. Thus, monitoring of levels of bioremediating bacteria becomes necessary in assessment of the return of the site to the natural state. Fungal bioremediation is also possible. This technology

utilizes white-rot fungi to clean up a wide spectrum of soil pollutants, such as wood preservatives, polycyclic aromatic hydrocarbons, organochlorines, polychlorinated biphyenyls, dyes, pesticides, fungicides, herbicides, and others. Rapid throughput methods for detection and identification of the members of the indigenous and added bacterial and/or fungal consortium

5 would greatly facilitate characterization of such bioremediative processes.

Household mold, is a growing concern for homeowners. Molds not only pose serious threats to a home's construction and survival, they pose serious threats to one's health. More than 100,000 types of molds have been discovered yet little is known about the life of these molds and their allergens, and how they become airborne.

10 *Stachybotrys* is a species of mold which has earned the title "toxic black mold," as it is one of the most lethal, yet common forms. *Stachybotrys* can become airborne and cause serious respiratory difficulties, memory and hearing loss, hemorrhaging, dizziness and sometimes death. Prolonged exposure to this strain can impair memory.

Cladosporium, Penicillium and Alternaria are more frequently detected in household

15 mold problems. While not as likely to pose a fatal threat, these molds are known for causing asthma-related symptoms. Studies suggest that such molds are culpable for, or at least connected to, the tripled asthma rate in the past 20 years (see, for example, www.paloaltoonline.com/paw/paonline/news_features/real_estate/spring2002/2002_03_13.mold. shtml).

20 Identification of species of molds would also benefit from a rapid method for detection and identification.

Mass spectrometry provides detailed information about the molecules being analyzed, including high mass accuracy. It is also a process that can be easily automated. Low-resolution MS may be unreliable when used to detect some known agents, if their spectral lines are

25 sufficiently weak or sufficiently close to those from other living organisms in the sample. DNA chips with specific probes can only determine the presence or absence of specifically anticipated organisms. Because there are hundreds of thousands of species of benign bacteria, some very similar in sequence to threat organisms, even arrays with 10,000 probes lack the breadth needed to detect a particular organism.

30 Antibodies face more severe diversity limitations than arrays. If antibodies are designed against highly conserved targets to increase diversity, the false alarm problem will dominate, again because threat organisms are very similar to benign ones. Antibodies are only capable of detecting known agents in relatively uncluttered environments.

-5-

Several groups have described detection of PCR products using high resolution electrospray ionization-Fourier transform-ion cyclotron resonance mass spectrometry (ESI-FT-ICR MS). Accurate measurement of exact mass combined with knowledge of the number of at least one nucleotide allowed calculation of the total base composition for PCR duplex products
5   of approximately 100 base pairs. (Aaserud *et al.*, *J. Am. Soc. Mass Spec.*, 1996, *7*, 1266-1269; Muddiman *et al.*, *Anal. Chem.*, 1997, *69*, 1543-1549; Wunschel *et al.*, *Anal. Chem.*, 1998, *70*, 1203-1207; Muddiman *et al.*, *Rev. Anal. Chem.*, 1998, *17*, 1-68). Electrospray ionization-Fourier transform-ion cyclotron resistance (ESI-FT-ICR) MS may be used to determine the mass of double-stranded, 500 base-pair PCR products via the average molecular mass (Hurst *et al.*, *Rapid*
10 *Commun. Mass Spec.* 1996, *10*, 377-382). The use of matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry for characterization of PCR products has been described. (Muddiman *et al.*, *Rapid Commun. Mass Spec.*, 1999, *13*, 1201-1204). However, the degradation of DNAs over about 75 nucleotides observed with MALDI limited the utility of this method.

15   U.S. Patent No. 5,849,492 describes a method for retrieval of phylogenetically informative DNA sequences which comprise searching for a highly divergent segment of genomic DNA surrounded by two highly conserved segments, designing the universal primers for PCR amplification of the highly divergent region, amplifying the genomic DNA by PCR technique using universal primers, and then sequencing the gene to determine the identity of the
20 organism.

U.S. Patent No. 5,965,363 discloses methods for screening nucleic acids for polymorphisms by analyzing amplified target nucleic acids using mass spectrometric techniques and to procedures for improving mass resolution and mass accuracy of these methods.

WO 99/14375 describes methods, PCR primers and kits for use in analyzing preselected
25 DNA tandem nucleotide repeat alleles by mass spectrometry.

WO 98/12355 discloses methods of determining the mass of a target nucleic acid by mass spectrometric analysis, by cleaving the target nucleic acid to reduce its length, making the target single-stranded and using MS to determine the mass of the single-stranded shortened target. Also disclosed are methods of preparing a double-stranded target nucleic acid for MS
30 analysis comprising amplification of the target nucleic acid, binding one of the strands to a solid support, releasing the second strand and then releasing the first strand which is then analyzed by MS. Kits for target nucleic acid preparation are also provided.

PCT WO97/33000 discloses methods for detecting mutations in a target nucleic acid by nonrandomly fragmenting the target into a set of single-stranded nonrandom length fragments and determining their masses by MS.

U.S. Patent No. 5,605,798 describes a fast and highly accurate mass spectrometer-based 5 process for detecting the presence of a particular nucleic acid in a biological sample for diagnostic purposes.

WO 98/21066 describes processes for determining the sequence of a particular target nucleic acid by mass spectrometry. Processes for detecting a target nucleic acid present in a biological sample by PCR amplification and mass spectrometry detection are disclosed, as are 10 methods for detecting a target nucleic acid in a sample by amplifying the target with primers that contain restriction sites and tags, extending and cleaving the amplified nucleic acid, and detecting the presence of extended product, wherein the presence of a DNA fragment of a mass different from wild-type is indicative of a mutation. Methods of sequencing a nucleic acid via mass spectrometry methods are also described.

15 WO 97/37041, WO 99/31278 and U.S. Patent No. 5,547,835 describe methods of sequencing nucleic acids using mass spectrometry. U.S. Patent Nos. 5,622,824, 5,872,003 and 5,691,141 describe methods, systems and kits for exonuclease-mediated mass spectrometric sequencing.

Thus, there is a need for a method for bioagent detection and identification which is 20 both specific and rapid, and in which no nucleic acid sequencing is required. The present invention addresses this need.

SUMMARY OF THE INVENTION

The present invention is directed to methods of identifying a bioagent in a sample by 25 determining a first molecular mass of a first amplification product of a first bioagent identifying amplicon from the sample and comparing the first molecular mass to a second molecular mass of a second bioagent identifying amplicon wherein both first and second bioagent identifying amplicons are correlative. These methods are applicable to environmental samples including, for example, air samples, water samples, soil samples, surface swab samples and samples from a 30 building or a container. These methods are also applicable to product samples including, for example, foodstuff and cosmetic samples.

The present invention is also directed to methods of monitoring a bioremediation process by identifying bioagents in a sample by determining a first molecular mass of a first amplification product of a first bioagent identifying amplicon from the sample and comparing

-7-

the first molecular mass to a second molecular mass of a second bioagent identifying amplicon wherein both first and second bioagent identifying amplicons are correlative.

BRIEF DESCRIPTION OF THE DRAWINGS

5   Figures 1A-1H and Figure 2 are consensus diagrams that show examples of conserved regions from 16S rRNA (Fig. 1A-1, 1A-2, 1A-3, 1A-4, and 1A-5), 23S rRNA (3'-half, Fig. 1B, 1C, and 1D; 5'-half, Fig. 1E-F), 23S rRNA Domain I (Fig. 1G), 23S rRNA Domain IV (Fig. 1H) and 16S rRNA Domain III (Fig. 2) which are suitable for use in the present invention. Lines with arrows are examples of regions to which intelligent primer pairs for PCR are designed. The label

10   for each primer pair represents the starting and ending base number of the amplified region on the consensus diagram. Bases in capital letters are greater than 95% conserved; bases in lower case letters are 90-95% conserved, filled circles are 80-90% conserved; and open circles are less than 80% conserved. The label for each primer pair represents the starting and ending base number of the amplified region on the consensus diagram. The nucleotide sequence of the 16S

15   rRNA consensus sequence is SEQ ID NO:3 and the nucleotide sequence of the 23S rRNA consensus sequence is SEQ ID NO:4.

        Figure 2 shows a typical primer amplified region from the 16S rRNA Domain III shown in Figure 1A-1.

        Figure 3 is a schematic diagram showing conserved regions in RNase P. Bases in capital

20   letters are greater than 90% conserved; bases in lower case letters are 80-90% conserved; filled circles designate bases which are 70-80% conserved; and open circles designate bases that are less than 70% conserved.

        Figure 4 is a schematic diagram of base composition signature determination using nucleotide analog "tags" to determine base composition signatures.

25   Figure 5 shows the deconvoluted mass spectra of a *Bacillus anthracis* region with and without the mass tag phosphorothioate A (A*). The two spectra differ in that the measured molecular weight of the mass tag-containing sequence is greater than the unmodified sequence.

        Figure 6 shows base composition signature (BCS) spectra from PCR products from *Staphylococcus aureus* (*S. aureus* 16S_1337F) and *Bacillus anthracis* (*B. anthr.* 16S_1337F),

30   amplified using the same primers. The two strands differ by only two (AT-->CG) substitutions and are clearly distinguished on the basis of their BCS.

        Figure 7 shows that a single difference between two sequences (A14 in *B. anthracis* vs. A15 in *B. cereus*) can be easily detected using ESI-TOF mass spectrometry.

Figure 8 is an ESI-TOF of *Bacillus anthracis* spore coat protein sspE 56mer plus calibrant. The signals unambiguously identify *B. anthracis* versus other Bacillus species.

Figure 9 is an ESI-TOF of a *B. anthracis* synthetic 16S_1228 duplex (reverse and forward strands). The technique easily distinguishes between the forward and reverse strands.

Figure 10 is an ESI-FTICR-MS of a synthetic *B. anthracis* 16S_1337 46 base pair duplex.

Figure 11 is an ESI-TOF-MS of a 56mer oligonucleotide (3 scans) from the *B. anthracis* saspB gene with an internal mass standard. The internal mass standards are designated by asterisks.

Figure 12 is an ESI-TOF-MS of an internal standard with 5 mM TBA-TFA buffer showing that charge stripping with tributylammonium trifluoroacetate reduces the most abundant charge state from [M-8H+]8- to [M-3H+]3-.

Figure 13 is a portion of a secondary structure defining database according to one embodiment of the present invention, where two examples of selected sequences are displayed graphically thereunder.

Figure 14 is a three dimensional graph demonstrating the grouping of sample molecular weight according to species.

Figure 15 is a three dimensional graph demonstrating the grouping of sample molecular weights according to species of virus and mammal infected.

Figure 16 is a three dimensional graph demonstrating the grouping of sample molecular weights according to species of virus, and animal-origin of infectious agent.

Figure 17 is a figure depicting how the triangulation method of the present invention provides for the identification of an unknown bioagent without prior knowledge of the unknown agent. The use of different primer sets to distinguish and identify the unknown is also depicted as primer sets I, II and III within this figure. A three dimensional graph depicts all of bioagent space (**170**), including the unknown bioagent, which after use of primer set I (**171**) according to a method according to the present invention further differentiates and classifies bioagents according to major classifications (**176**) which, upon further analysis using primer set II (**172**) differentiates the unknown agent (**177**) from other, known agents (**173**) and finally, the use of a third primer set (**175**) further specifies subgroups within the family of the unknown (**174**).

Figure 18 shows a representative mass spectra from amplicons derived from a single primer pair used in a dilution to extinction experiment with Bti spores as described. The concentration of spores is shown at the right of each spectrum. Positive and negative strands are labeled.

Figure 19 shows a representative ROC curve demonstrating that air samples spiked with a spore sample are identified and detectable above background bioagents by using the present methods.

5 **DESCRIPTION OF EMBODIMENTS**

The present invention provides, *inter alia*, methods for detection and identification of bioagents in an unbiased manner using "bioagent identifying amplicons." "Intelligent primers" are selected to hybridize to conserved sequence regions of nucleic acids derived from a bioagent and which bracket variable sequence regions to yield a bioagent identifying amplicon which can

10 be amplified and which is amenable to molecular mass determination. The molecular mass then provides a means to uniquely identify the bioagent without a requirement for prior knowledge of the possible identity of the bioagent. The molecular mass or corresponding "base composition signature" (BCS) of the amplification product is then matched against a database of molecular masses or base composition signatures. Furthermore, the method can be applied to rapid parallel

15 "multiplex" analyses, the results of which can be employed in a triangulation identification strategy. The present method provides rapid throughput and does not require nucleic acid sequencing of the amplified target sequence for bioagent detection and identification.

In the context of this invention, a "bioagent" is any organism, cell, or virus, living or dead, or a nucleic acid derived from such an organism, cell or virus. Examples of bioagents

20 include, but are not limited to, cells, including but not limited to, cells, including but not limited to human clinical samples, bacterial cells and other pathogens) viruses, fungi, and protists, parasites, and pathogenicity markers (including but not limited to: pathogenicity islands, antibiotic resistance genes, virulence factors, toxin genes and other bioregulating compounds). Samples may be alive or dead or in a vegetative state (for example, vegetative bacteria or spores)

25 and may be encapsulated or bioengineered. In the context of this invention, a "pathogen" is a bioagent which causes a disease or disorder.

Despite enormous biological diversity, all forms of life on earth share sets of essential, common features in their genomes. Bacteria, for example have highly conserved sequences in a variety of locations on their genomes. Most notable is the universally conserved region of the

30 ribosome. but there are also conserved elements in other non-coding RNAs, including RNAse P and the signal recognition particle (SRP) among others. Bacteria have a common set of absolutely required genes. About 250 genes are present in all bacterial species (*Proc. Natl. Acad. Sci. U.S.A.*, **1996**, *93*, 10268; *Science*, **1995**, *270*, 397), including tiny genomes like *Mycoplasma*, *Ureaplasma* and *Rickettsia*. These genes encode proteins involved in translation, replication,

recombination and repair, transcription, nucleotide metabolism, amino acid metabolism, lipid metabolism, energy generation, uptake, secretion and the like. Examples of these proteins are DNA polymerase III beta, elongation factor TU, heat shock protein groEL, RNA polymerase beta, phosphoglycerate kinase, NADH dehydrogenase, DNA ligase, DNA topoisomerase and

5 elongation factor G. Operons can also be targeted using the present method. One example of an operon is the bfp operon from enteropathogenic *E. coli*. Multiple core chromosomal genes can be used to classify bacteria at a genus or genus species level to determine if an organism has threat potential. The methods can also be used to detect pathogenicity markers (plasmid or chromosomal) and antibiotic resistance genes to confirm the threat potential of an organism and

10 to direct countermeasures.

Since genetic data provide the underlying basis for identification of bioagents by the methods of the present invention, it is necessary to select segments of nucleic acids which ideally provide enough variability to distinguish each individual bioagent and whose molecular mass is amenable to molecular mass determination. In one embodiment of the present invention, at least

15 one polynucleotide segment is amplified to facilitate detection and analysis in the process of identifying the bioagent. Thus, the nucleic acid segments which provide enough variability to distinguish each individual bioagent and whose molecular masses are amenable to molecular mass determination are herein described as "bioagent identifying amplicons." The term "amplicon" as used herein, refers to a segment of a polynucleotide which is amplified in an

20 amplification reaction.

As used herein, "intelligent primers" are primers that are designed to bind to highly conserved sequence regions of a bioagent identifying amplicon that flank an intervening variable region and yield amplification products which ideally provide enough variability to distinguish each individual bioagent, and which are amenable to molecular mass analysis. By the term

25 "highly conserved," it is meant that the sequence regions exhibit between about 80-100%, or between about 90-100%, or between about 95-100% identity. The molecular mass of a given amplification product provides a means of identifying the bioagent from which it was obtained, due to the variability of the variable region. Thus design of intelligent primers requires selection of a variable region with appropriate variability to resolve the identity of a given bioagent.

30 Bioagent identifying amplicons are ideally specific to the identity of the bioagent. A plurality of bioagent identifying amplicons selected in parallel for distinct bioagents which contain the same conserved sequences for hybridization of the same pair of intelligent primers are herein defined as "correlative bioagent identifying amplicons."

In one embodiment, the bioagent identifying amplicon is a portion of a ribosomal RNA (rRNA) gene sequence. With the complete sequences of many of the smallest microbial genomes now available, it is possible to identify a set of genes that defines "minimal life" and identify composition signatures that uniquely identify each gene and organism. Genes that encode core

5   life functions such as DNA replication, transcription, ribosome structure, translation, and transport are distributed broadly in the bacterial genome and are suitable regions for selection of bioagent identifying amplicons. Ribosomal RNA (rRNA) genes comprise regions that provide useful base composition signatures. Like many genes involved in core life functions, rRNA genes contain sequences that are extraordinarily conserved across bacterial domains interspersed

10 with regions of high variability that are more specific to each species. The variable regions can be utilized to build a database of base composition signatures. The strategy involves creating a structure-based alignment of sequences of the small (16S) and the large (23S) subunits of the rRNA genes. For example, there are currently over 13,000 sequences in the ribosomal RNA database that has been created and maintained by Robin Gutell, University of Texas at Austin,

15 and is publicly available on the Institute for Cellular and Molecular Biology web page on the world wide web of the Internet at, for example, "rna.icmb.utexas.edu/." There is also a publicly available rRNA database created and maintained by the University of Antwerp, Belgium on the world wide web of the Internet at, for example, "rrna.uia.ac.be."

These databases have been analyzed to determine regions that are useful as bioagent

20 identifying amplicons. The characteristics of such regions include: a) between about 80 and 100%, or greater than about 95% identity among species of the particular bioagent of interest, of upstream and downstream nucleotide sequences which serve as sequence amplification primer sites; b) an intervening variable region which exhibits no greater than about 5% identity among species; and c) a separation of between about 30 and 1000 nucleotides, or no more than about

25 50-250 nucleotides, or no more than about 60-100 nucleotides, between the conserved regions.

As a non-limiting example, for identification of *Bacillus* species, the conserved sequence regions of the chosen bioagent identifying amplicon must be highly conserved among all *Bacillus* species while the variable region of the bioagent identifying amplicon is sufficiently variable such that the molecular masses of the amplification products of all species of *Bacillus*

30 are distinguishable.

Bioagent identifying amplicons amenable to molecular mass determination are either of a length, size or mass compatible with the particular mode of molecular mass determination or compatible with a means of providing a predictable fragmentation pattern in order to obtain predictable fragments of a length compatible with the particular mode of molecular mass

determination. Such means of providing a predictable fragmentation pattern of an amplification product include, but are not limited to, cleavage with restriction enzymes or cleavage primers, for example.

Identification of bioagents can be accomplished at different levels using intelligent
5   primers suited to resolution of each individual level of identification. "Broad range survey" intelligent primers are designed with the objective of identifying a bioagent as a member of a particular division of bioagents. A "bioagent division" is defined as group of bioagents above the species level and includes but is not limited to: orders, families, classes, clades, genera or other such groupings of bioagents above the species level. As a non-limiting example, members of the
10 *Bacillus/Clostridia* group or gamma-proteobacteria group may be identified as such by employing broad range survey intelligent primers such as primers which target 16S or 23S ribosomal RNA.

In some embodiments, broad range survey intelligent primers are capable of identification of bioagents at the species level. One main advantage of the detection methods of
15 the present invention is that the broad range survey intelligent primers need not be specific for a particular bacterial species, or even genus, such as *Bacillus* or *Streptomyces*. Instead, the primers recognize highly conserved regions across hundreds of bacterial species including, but not limited to, the species described herein. Thus, the same broad range survey intelligent primer pair can be used to identify any desired bacterium because it will bind to the conserved regions
20 that flank a variable region specific to a single species, or common to several bacterial species, allowing unbiased nucleic acid amplification of the intervening sequence and determination of its molecular weight and base composition. For example, the 16S_971-1062, 16S_1228-1310 and 16S_1100-1188 regions are 98-99% conserved in about 900 species of bacteria (16S=16S rRNA, numbers indicate nucleotide position). In one embodiment of the present invention, primers used
25 in the present method bind to one or more of these regions or portions thereof.

Due to their overall conservation, the flanking rRNA primer sequences serve as good intelligent primer binding sites to amplify the nucleic acid region of interest for most, if not all, bacterial species. The intervening region between the sets of primers varies in length and/or composition, and thus provides a unique base composition signature. Examples of intelligent
30 primers that amplify regions of the 16S and 23S rRNA are shown in Figures 1A-1H. A typical primer amplified region in 16S rRNA is shown in Figure 2. The arrows represent primers that bind to highly conserved regions which flank a variable region in 16S rRNA domain III. The amplified region is the stem-loop structure under "1100-1188." It is advantageous to design the broad range survey intelligent primers to minimize the number of primers required for the

-13-

analysis, and to allow detection of multiple members of a bioagent division using a single pair of primers. The advantage of using broad range survey intelligent primers is that once a bioagent is broadly identified, the process of further identification at species and sub-species levels is facilitated by directing the choice of additional intelligent primers.

5   "Division-wide" intelligent primers are designed with an objective of identifying a bioagent at the species level. As a non-limiting example, a *Bacillus anthracis, Bacillus cereus* and *Bacillus thuringiensis* can be distinguished from each other using division-wide intelligent primers. Division-wide intelligent primers are not always required for identification at the species level because broad range survey intelligent primers may provide sufficient identification
10   resolution to accomplishing this identification objective.

         "Drill-down" intelligent primers are designed with an objective of identifying a sub-species characteristic of a bioagent. A "sub-species characteristic" is defined as a property imparted to a bioagent at the sub-species level of identification as a result of the presence or absence of a particular segment of nucleic acid. Such sub-species characteristics include, but are
15   not limited to, strains, sub-types, pathogenicity markers such as antibiotic resistance genes, pathogenicity islands, toxin genes and virulence factors. Identification of such sub-species characteristics is often critical for determining proper clinical treatment of pathogen infections.
Chemical Modifications of Intelligent Primers

         Ideally, intelligent primer hybridization sites are highly conserved in order to facilitate
20   the hybridization of the primer. In cases where primer hybridization is less efficient due to lower levels of conservation of sequence, intelligent primers can be chemically modified to improve the efficiency of hybridization.

         For example, because any variation (due to codon wobble in the 3rd position) in these conserved regions among species is likely to occur in the third position of a DNA triplet,
25   oligonucleotide primers can be designed such that the nucleotide corresponding to this position is a base which can bind to more than one nucleotide, referred to herein as a "universal base." For example, under this "wobble" pairing, inosine (I) binds to U, C or A; guanine (G) binds to U or C, and uridine (U) binds to U or C. Other examples of universal bases include nitroindoles such as 5-nitroindole or 3-nitropyrrole (Loakes *et al., Nucleosides and Nucleotides*, **1995**, *14*, 1001-
30   1003), the degenerate nucleotides dP or dK (Hill *et al.*), an acyclic nucleoside analog containing 5-nitroindazole (Van Aerschot *et al., Nucleosides and Nucleotides*, **1995**, *14*, 1053-1056) or the purine analog 1-(2-deoxy-β-D-ribofuranosyl)-imidazole-4-carboxamide (Sala *et al., Nucl. Acids Res.*, **1996**, *24*, 3302-3306).

In another embodiment of the invention, to compensate for the somewhat weaker binding by the "wobble" base, the oligonucleotide primers are designed such that the first and second positions of each triplet are occupied by nucleotide analogs which bind with greater affinity than the unmodified nucleotide. Examples of these analogs include, but are not limited to, 2,6-diaminopurine which binds to thymine, propyne T which binds to adenine and propyne C and phenoxazines, including G-clamp, which binds to G. Propynylated pyrimidines are described in U.S. Patent Nos. 5,645,985, 5,830,653 and 5,484,908, each of which is commonly owned and incorporated herein by reference in its entirety. Propynylated primers are claimed in U.S Serial No. 10/294,203 which is also commonly owned and incorporated herein by reference in entirety. Phenoxazines are described in U.S. Patent Nos. 5,502,177, 5,763,588, and 6,005,096, each of which is incorporated herein by reference in its entirety. G-clamps are described in U.S. Patent Nos. 6,007,992 and 6,028,183, each of which is incorporated herein by reference in its entirety.

A theoretically ideal bioagent detector would identify, quantify, and report the complete nucleic acid sequence of every bioagent that reached the sensor. The complete sequence of the nucleic acid component of a pathogen would provide all relevant information about the threat, including its identity and the presence of drug-resistance or pathogenicity markers. This ideal has not yet been achieved. However, the present invention provides a straightforward strategy for obtaining information with the same practical value based on analysis of bioagent identifying amplicons by molecular mass determination.

In some cases, a molecular mass of a given bioagent identifying amplicon alone does not provide enough resolution to unambiguously identify a given bioagent. For example, the molecular mass of the bioagent identifying amplicon obtained using the intelligent primer pair "16S_971" would be 55622 Da for both *E. coli* and *Salmonella typhimurium*. However, if additional intelligent primers are employed to analyze additional bioagent identifying amplicons, a "triangulation identification" process is enabled. For example, the "16S_1100" intelligent primer pair yields molecular masses of 55009 and 55005 Da for *E. coli* and *Salmonella typhimurium,* respectively. Furthermore, the "23S_855" intelligent primer pair yields molecular masses of 42656 and 42698 Da for *E. coli* and *Salmonella typhimurium,* respectively. In this basic example, the second and third intelligent primer pairs provided the additional "fingerprinting" capability or resolution to distinguish between the two bioagents.

In another embodiment, the triangulation identification process is pursued by measuring signals from a plurality of bioagent identifying amplicons selected within multiple core genes. This process is used to reduce false negative and false positive signals, and enable reconstruction of the origin of hybrid or otherwise engineered bioagents. In this process, after identification of

multiple core genes, alignments are created from nucleic acid sequence databases. The
alignments are then analyzed for regions of conservation and variation, and bioagent identifying
amplicons are selected to distinguish bioagents based on specific genomic differences. For
example, identification of the three part toxin genes typical of *B. anthracis* (Bowen *et al.*, *J.*
5 *Appl. Microbiol.*, 1999, 87, 270-278) in the absence of the expected signatures from the *B.*
anthracis genome would suggest a genetic engineering event.

The triangulation identification process can be pursued by characterization of bioagent
identifying amplicons in a massively parallel fashion using the polymerase chain reaction (PCR),
such as multiplex PCR, and mass spectrometric (MS) methods. Sufficient quantities of nucleic
10 acids should be present for detection of bioagents by MS. A wide variety of techniques for
preparing large amounts of purified nucleic acids or fragments thereof are well known to those of
skill in the art. PCR requires one or more pairs of oligonucleotide primers that bind to regions
which flank the target sequence(s) to be amplified. These primers prime synthesis of a different
strand of DNA, with synthesis occurring in the direction of one primer towards the other primer.
15 The primers, DNA to be amplified, a thermostable DNA polymerase (e.g. *Taq* polymerase), the
four deoxynucleotide triphosphates, and a buffer are combined to initiate DNA synthesis. The
solution is denatured by heating, then cooled to allow annealing of newly added primer, followed
by another round of DNA synthesis. This process is typically repeated for about 30 cycles,
resulting in amplification of the target sequence.
20 Although the use of PCR is suitable, other nucleic acid amplification techniques may
also be used, including ligase chain reaction (LCR) and strand displacement amplification
(SDA). The high-resolution MS technique allows separation of bioagent spectral lines from
background spectral lines in highly cluttered environments.

In another embodiment, the detection scheme for the PCR products generated from the
25 bioagent(s) incorporates at least three features. First, the technique simultaneously detects and
differentiates multiple (generally about 6-10) PCR products. Second, the technique provides a
molecular mass that uniquely identifies the bioagent from the possible primer sites. Finally, the
detection technique is rapid, allowing multiple PCR reactions to be run in parallel.

Mass spectrometry (MS)-based detection of PCR products provides a means for
30 determination of BCS which has several advantages. MS is intrinsically a parallel detection
scheme without the need for radioactive or fluorescent labels, since every amplification product
is identified by its molecular mass. The current state of the art in mass spectrometry is such that
less than femtomole quantities of material can be readily analyzed to afford information about
the molecular contents of the sample. An accurate assessment of the molecular mass of the

material can be quickly obtained, irrespective of whether the molecular weight of the sample is several hundred, or in excess of one hundred thousand atomic mass units (amu) or Daltons. Intact molecular ions can be generated from amplification products using one of a variety of ionization techniques to convert the sample to gas phase. These ionization methods include, but

5 are not limited to, electrospray ionization (ES), matrix-assisted laser desorption ionization (MALDI) and fast atom bombardment (FAB). For example, MALDI of nucleic acids, along with examples of matrices for use in MALDI of nucleic acids, are described in WO 98/54751 (Genetrace, Inc.).

       In some embodiments, large DNAs and RNAs, or large amplification products

10 therefrom, can be digested with restriction endonucleases prior to ionization. Thus, for example, an amplification product that was 10 kDa could be digested with a series of restriction endonucleases to produce a panel of, for example, 100 Da fragments. Restriction endonucleases and their sites of action are well known to the skilled artisan. In this manner, mass spectrometry can be performed for the purposes of restriction mapping.

15   Upon ionization, several peaks are observed from one sample due to the formation of ions with different charges. Averaging the multiple readings of molecular mass obtained from a single mass spectrum affords an estimate of molecular mass of the bioagent. Electrospray ionization mass spectrometry (ESI-MS) is particularly useful for very high molecular weight polymers such as proteins and nucleic acids having molecular weights greater than 10 kDa, since

20 it yields a distribution of multiply-charged molecules of the sample without causing a significant amount of fragmentation.

       The mass detectors used in the methods of the present invention include, but are not limited to, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), ion trap, quadrupole, magnetic sector, time of flight (TOF), Q-TOF, and triple quadrupole.

25   In general, the mass spectrometric techniques which can be used in the present invention include, but are not limited to, tandem mass spectrometry, infrared multiphoton dissociation and pyrolytic gas chromatography mass spectrometry (PGC-MS). In one embodiment of the invention, the bioagent detection system operates continually in bioagent detection mode using pyrolytic GC-MS without PCR for rapid detection of increases in biomass

30 (for example, increases in fecal contamination of drinking water or of germ warfare agents). To achieve minimal latency, a continuous sample stream flows directly into the PGC-MS combustion chamber. When an increase in biomass is detected, a PCR process is automatically initiated. Bioagent presence produces elevated levels of large molecular fragments from, for example, about 100-7,000 Da which are observed in the PGC-MS spectrum. The observed mass

spectrum is compared to a threshold level and when levels of biomass are determined to exceed a predetermined threshold, the bioagent classification process described hereinabove (combining PCR and MS, such as FT-ICR MS) is initiated. Optionally, alarms or other processes (halting ventilation flow, physical isolation) are also initiated by this detected biomass level.

5   The accurate measurement of molecular mass for large DNAs is limited by the adduction of cations from the PCR reaction to each strand, resolution of the isotopic peaks from natural abundance ^{13}C and ^{15}N isotopes, and assignment of the charge state for any ion. The cations are removed by in-line dialysis using a flow-through chip that brings the solution containing the PCR products into contact with a solution containing ammonium acetate in the

10 presence of an electric field gradient orthogonal to the flow. The latter two problems are addressed by operating with a resolving power of >100,000 and by incorporating isotopically depleted nucleotide triphosphates into the DNA. The resolving power of the instrument is also a consideration. At a resolving power of 10,000, the modeled signal from the $[M-14H+]^{14-}$ charge state of an 84mer PCR product is poorly characterized and assignment of the charge state or

15 exact mass is impossible. At a resolving power of 33,000, the peaks from the individual isotopic components are visible. At a resolving power of 100,000, the isotopic peaks are resolved to the baseline and assignment of the charge state for the ion is straightforward. The $[^{13}C,^{15}N]$-depleted triphosphates are obtained, for example, by growing microorganisms on depleted media and harvesting the nucleotides (Batey *et al.*, *Nucl. Acids Res.*, **1992**, *20*, 4515-4523).

20   While mass measurements of intact nucleic acid regions are believed to be adequate to determine most bioagents, tandem mass spectrometry (MS^n) techniques may provide more definitive information pertaining to molecular identity or sequence. Tandem MS involves the coupled use of two or more stages of mass analysis where both the separation and detection steps are based on mass spectrometry. The first stage is used to select an ion or component of a sample

25 from which further structural information is to be obtained. The selected ion is then fragmented using, e.g., blackbody irradiation, infrared multiphoton dissociation, or collisional activation. For example, ions generated by electrospray ionization (ESI) can be fragmented using IR multiphoton dissociation. This activation leads to dissociation of glycosidic bonds and the phosphate backbone, producing two series of fragment ions, called the *w*-series (having an intact

30 3' terminus and a 5' phosphate following internal cleavage) and the *a*-Base series (having an intact 5' terminus and a 3' furan).

          The second stage of mass analysis is then used to detect and measure the mass of these resulting fragments of product ions. Such ion selection followed by fragmentation routines can

be performed multiple times so as to essentially completely dissect the molecular sequence of a
sample.

If there are two or more targets of similar molecular mass, or if a single amplification
reaction results in a product which has the same mass as two or more bioagent reference
5   standards, they can be distinguished by using mass-modifying "tags." In this embodiment of the
invention, a nucleotide analog or "tag" is incorporated during amplification (e.g., a 5-
(trifluoromethyl) deoxythymidine triphosphate) which has a different molecular weight than the
unmodified base so as to improve distinction of masses. Such tags are described in, for example,
PCT WO97/33000, which is incorporated herein by reference in its entirety. This further limits
10 the number of possible base compositions consistent with any mass. For example, 5-
(trifluoromethyl)deoxythymidine triphosphate can be used in place of dTTP in a separate nucleic
acid amplification reaction. Measurement of the mass shift between a conventional amplification
product and the tagged product is used to quantitate the number of thymidine nucleotides in each
of the single strands. Because the strands are complementary, the number of adenosine
15 nucleotides in each strand is also determined.

In another amplification reaction, the number of G and C residues in each strand is
determined using, for example, the cytidine analog 5-methylcytosine (5-meC) or propyne C. The
combination of the A/T reaction and G/C reaction, followed by molecular weight determination,
provides a unique base composition. This method is summarized in Figure 4 and Table 1.

20   **Table 1**

Mass tag	Double strand sequence	Single strand Sequence	Total mass this strand	Base info this strand	Base info other strand	Total base comp. Top strand	Total base comp. Bottom strand
T*mass (T*-T) = x	T*ACGT*ACGT * AT*GCAT*GCA	T*ACGT*ACGT*	3x	3T	3A	3T 2A 2C 2G	3A 2T 2G 2C
		AT*GCAT*GCA	2x	2T	2A		
C*mass (C*-C) = y	TAC*GTAC*GT ATGC*ATGC*A	TAC*GTAC*GT	2x	2C	2G		
		ATGC*ATGC*A	2x	2C	2G		

-19-

The mass tag phosphorothioate A (A*) was used to distinguish a *Bacillus anthracis* cluster. The *B. anthracis* ($A_{14}G_9C_{14}T_9$) had an average MW of 14072.26, and the *B. anthracis* ($A_1A*_{13}G_9C_{14}T_9$) had an average molecular weight of 14281.11 and the phosphorothioate A had an average molecular weight of +16.06 as determined by ESI-TOF MS. The deconvoluted

5   spectra are shown in Figure 5.

In another example, assume the measured molecular masses of each strand are 30,000.115Da and 31,000.115 Da respectively, and the measured number of dT and dA residues are (30,28) and (28,30). If the molecular mass is accurate to 100 ppm, there are 7 possible combinations of dG+dC possible for each strand. However, if the measured molecular mass is

10   accurate to 10 ppm, there are only 2 combinations of dG+dC, and at 1 ppm accuracy there is only one possible base composition for each strand.

Signals from the mass spectrometer may be input to a maximum-likelihood detection and classification algorithm such as is widely used in radar signal processing. The detection processing uses matched filtering of BCS observed in mass-basecount space and allows for

15   detection and subtraction of signatures from known, harmless organisms, and for detection of unknown bioagent threats. Comparison of newly observed bioagents to known bioagents is also possible, for estimation of threat level, by comparing their BCS to those of known organisms and to known forms of pathogenicity enhancement, such as insertion of antibiotic resistance genes or toxin genes.

20   Processing may end with a Bayesian classifier using log likelihood ratios developed from the observed signals and average background levels. The program emphasizes performance predictions culminating in probability-of-detection versus probability-of-false-alarm plots for conditions involving complex backgrounds of naturally occurring organisms and environmental contaminants. Matched filters consist of a priori expectations of signal values given the set of

25   primers used for each of the bioagents. A genomic sequence database (e.g. GenBank) is used to define the mass basecount matched filters. The database contains known threat agents and benign background organisms. The latter is used to estimate and subtract the signature produced by the background organisms. A maximum likelihood detection of known background organisms is implemented using matched filters and a running-sum estimate of the noise covariance.

30   Background signal strengths are estimated and used along with the matched filters to form signatures which are then subtracted. the maximum likelihood process is applied to this "cleaned up" data in a similar manner employing matched filters for the organisms and a running-sum estimate of the noise-covariance for the cleaned up data.

Although the molecular mass of amplification products obtained using intelligent primers provides a means for identification of bioagents, conversion of molecular mass data to a base composition signature is useful for certain analyses. As used herein, a "base composition signature" (BCS) is the exact base composition determined from the molecular mass of a

5   bioagent identifying amplicon. In one embodiment, a BCS provides an index of a specific gene in a specific organism.

Base compositions, like sequences, vary slightly from isolate to isolate within species. It is possible to manage this diversity by building "base composition probability clouds" around the composition constraints for each species. This permits identification of organisms in a fashion

10   similar to sequence analysis. A "pseudo four-dimensional plot" can be used to visualize the concept of base composition probability clouds (Figure 18). Optimal primer design requires optimal choice of bioagent identifying amplicons and maximizes the separation between the base composition signatures of individual bioagents. Areas where clouds overlap indicate regions that may result in a misclassification, a problem which is overcome by selecting primers that provide

15   information from different bioagent identifying amplicons, ideally maximizing the separation of base compositions. Thus, one aspect of the utility of an analysis of base composition probability clouds is that it provides a means for screening primer sets in order to avoid potential misclassifications of BCS and bioagent identity. Another aspect of the utility of base composition probability clouds is that they provide a means for predicting the identity of a

20   bioagent whose exact measured BCS was not previously observed and/or indexed in a BCS database due to evolutionary transitions in its nucleic acid sequence.

It is important to note that, in contrast to probe-based techniques, mass spectrometry determination of base composition does not require prior knowledge of the composition in order to make the measurement, only to interpret the results. In this regard, the present invention

25   provides bioagent classifying information similar to DNA sequencing and phylogenetic analysis at a level sufficient to detect and identify a given bioagent. Furthermore, the process of determination of a previously unknown BCS for a given bioagent (for example, in a case where sequence information is unavailable) has downstream utility by providing additional bioagent indexing information with which to populate BCS databases. The process of future bioagent

30   identification is thus greatly improved as more BCS indexes become available in the BCS databases.

Another embodiment of the present invention is a method of surveying bioagent samples that enables detection and identification of all bacteria for which sequence information is available using a set of twelve broad-range intelligent PCR primers. Six of the twelve primers

are "broad range survey primers" herein defined as primers targeted to broad divisions of bacteria (for example, the *Bacillus/Clostridia* group or gamma-proteobacteria). The other six primers of the group of twelve primers are "division-wide" primers herein defined as primers which provide more focused coverage and higher resolution. This method enables identification of nearly 100% of known bacteria at the species level. A further example of this embodiment of the present invention is a method herein designated "survey/drill-down" wherein a subspecies characteristic for detected bioagents is obtained using additional primers. Examples of such a subspecies characteristic include but are not limited to: antibiotic resistance, pathogenicity island, virulence factor, strain type, sub-species type, and clade group. Using the survey/drill-down method, bioagent detection, confirmation and a subspecies characteristic can be provided within hours. Moreover, the survey/drill-down method can be focused to identify bioengineering events such as the insertion of a toxin gene into a bacterial species that does not normally make the toxin.

The present methods allow extremely rapid and accurate detection and identification of bioagents compared to existing methods. Furthermore, this rapid detection and identification is possible even when sample material is impure. The methods leverage ongoing biomedical research in virulence, pathogenicity, drug resistance and genome sequencing into a method which provides greatly improved sensitivity, specificity and reliability compared to existing methods, with lower rates of false positives. Thus, the methods are useful in a wide variety of fields, including, but not limited to, those fields discussed below.

In some embodiments of the invention, the methods disclosed herein can be used for environmental testing. "Environment" is herein defined as including both natural environment such as soil, water, living matter such as plants, as well as environments created by humans such as buildings, vehicles, containers, water towers. Detection and discrimination of pathogenic vs. non-pathogenic bacteria, viruses, parasites, fungi and the like, in samples of water, land, air, or other samples, can be carried out. Water samples can be obtained from, for example, lakes, rivers, oceans, streams, water treatment systems, rainwater, groundwater, water table, reservoirs, wells, bottled water, and the like. Air samples can be obtained from ventilation systems, airplane cabins, schools, hospitals, mass transit locations such as subways, train stations, airports, and the like. Land samples can be obtained from any location.

In other embodiments of the invention, the methods disclosed herein can be used for detecting the presence of bioagents in a container, such as a package, box, envelope, mail tube, railroad box car, and the like. For example, mail and package delivery entities and agencies, both domestic and abroad, as well investigative agencies such as the FBI and ATF can use the present

methods to detect bioagents in containers. Appropriate sampling techniques are well known to those skilled in the art.

In other embodiments of the invention, the methods disclosed herein can be used for detecting the presence of pathogenic and non-pathogenic bacteria, viruses, parasites, fungi and

5 the like in products. "Products" are defined as objects for consumption such as processed food, drinks and cosmetics. For example, food and wine can be examined for the presence of pathogenic and non-pathogenic bacteria, viruses, parasites, fungi and the like. Particular types of foods susceptible to bioagent contamination, such as agricultural products, meat products and eggs, can be examined for pathogenic organisms such as *E. coli* and *Salmonella* species. Such

10 examination procedures can be used by, for example, the wholesalers of foodstuffs and beverages, or by regulatory agencies such as the U.S. Department of Agriculture and the Food and Drug Administration. In addition, grapes and wines, for example, can be examined using the present methods to detect particular strains of bacteria or yeast that may indicate a particular time upon which to harvest the grapes or alter the wine-making process. Appropriate methods of

15 sampling food, drink and cosmetic products are well known to those will skill in the art.

In another embodiment, the present invention can be used for rapid detection and identification of species of household mold which is becoming a growing concern for homeowners. Methods of sampling household molds are well known to those skilled in the art.

In another embodiment, the methods of the present invention can be used to monitor

20 bioremediation processes by detection, identification and quantification of indigenous and bioremediating bioagents. Methods of sampling sites of bioremediation include, but are not limited to, water and soil sampling methods which are well known to those skilled in the art.

While the present invention has been described with specificity in accordance with certain of its embodiments, the following examples serve only to illustrate the invention and are

25 not intended to limit the same.

EXAMPLES
Example 1: Nucleic Acid Isolation and PCR

In one embodiment, nucleic acid is isolated from the organisms and amplified by PCR

30 using standard methods prior to BCS determination by mass spectrometry. Nucleic acid is isolated, for example, by detergent lysis of bacterial cells, centrifugation and ethanol precipitation. Nucleic acid isolation methods are described in, for example, *Current Protocols in Molecular Biology* (Ausubel et al.) and *Molecular Cloning; A Laboratory Manual* (Sambrook *et al.*). The nucleic acid is then amplified using standard methodology, such as PCR, with primers

-23-

which bind to conserved regions of the nucleic acid which contain an intervening variable sequence as described below.

General Genomic DNA Sample Prep Protocol: Raw samples are filtered using Supor-200 0.2 μm membrane syringe filters (VWR International) . Samples are transferred to 1.5 ml eppendorf tubes pre-filled with 0.45 g of 0.7 mm Zirconia beads followed by the addition of 350 μl of ATL buffer (Qiagen, Valencia, CA). The samples are subjected to bead beating for 10 minutes at a frequency of 19 l/s in a Retsch Vibration Mill (Retsch). After centrifugation, samples are transferred to an S-block plate (Qiagen) and DNA isolation is completed with a BioRobot 8000 nucleic acid isolation robot (Qiagen).

Swab Sample Protocol: Allegiance S/P brand culture swabs and collection/transport system are used to collect samples. After drying, swabs are placed in 17x100 mm culture tubes (VWR International) and the genomic nucleic acid isolation is carried out automatically with a Qiagen Mdx robot and the Qiagen QIAamp DNA Blood BioRobot Mdx genomic preparation kit (Qiagen, Valencia, CA).

Example 2: Mass spectrometry

FTICR Instrumentation: The FTICR instrument is based on a 7 tesla actively shielded superconducting magnet and modified Bruker Daltonics Apex II 70e ion optics and vacuum chamber. The spectrometer is interfaced to a LEAP PAL autosampler and a custom fluidics control system for high throughput screening applications. Samples are analyzed directly from 96-well or 384-well microtiter plates at a rate of about 1 sample/minute. The Bruker data-acquisition platform is supplemented with a lab-built ancillary NT datastation which controls the autosampler and contains an arbitrary waveform generator capable of generating complex rf-excite waveforms (frequency sweeps, filtered noise, stored waveform inverse Fourier transform (SWIFT), etc.) for sophisticated tandem MS experiments. For oligonucleotides in the 20-30-mer regime typical performance characteristics include mass resolving power in excess of 100,000 (FWHM), low ppm mass measurement errors, and an operable *m/z* range between 50 and 5000 *m/z*.

Modified ESI Source: In sample-limited analyses, analyte solutions are delivered at 150 nL/minute to a 30 mm i.d. fused-silica ESI emitter mounted on a 3-D micromanipulator. The ESI ion optics consists of a heated metal capillary, an rf-only hexapole, a skimmer cone, and an auxiliary gate electrode. The 6.2 cm rf-only hexapole is comprised of 1 mm diameter rods and is operated at a voltage of 380 Vpp at a frequency of 5 MHz. A lab-built electro-mechanical shutter can be employed to prevent the electrospray plume from entering the inlet capillary unless

triggered to the "open" position via a TTL pulse from the data station. When in the "closed" position, a stable electrospray plume is maintained between the ESI emitter and the face of the shutter. The back face of the shutter arm contains an elastomeric seal that can be positioned to form a vacuum seal with the inlet capillary. When the seal is removed, a 1 mm gap between the

5 shutter blade and the capillary inlet allows constant pressure in the external ion reservoir regardless of whether the shutter is in the open or closed position. When the shutter is triggered, a "time slice" of ions is allowed to enter the inlet capillary and is subsequently accumulated in the external ion reservoir. The rapid response time of the ion shutter (< 25 ms) provides reproducible, user defined intervals during which ions can be injected into and accumulated in

10 the external ion reservoir.

   Apparatus for Infrared Multiphoton Dissociation: A 25 watt CW CO_2 laser operating at 10.6 μm has been interfaced to the spectrometer to enable infrared multiphoton dissociation (IRMPD) for oligonucleotide sequencing and other tandem MS applications. An aluminum optical bench is positioned approximately 1.5 m from the actively shielded superconducting

15 magnet such that the laser beam is aligned with the central axis of the magnet. Using standard IR-compatible mirrors and kinematic mirror mounts, the unfocused 3 mm laser beam is aligned to traverse directly through the 3.5 mm holes in the trapping electrodes of the FTICR trapped ion cell and longitudinally traverse the hexapole region of the external ion guide finally impinging on the skimmer cone. This scheme allows IRMPD to be conducted in an *m/z* selective manner in

20 the trapped ion cell (e.g. following a SWIFT isolation of the species of interest), or in a broadband mode in the high pressure region of the external ion reservoir where collisions with neutral molecules stabilize IRMPD-generated metastable fragment ions resulting in increased fragment ion yield and sequence coverage.

25 **Example 3: Identification of Bioagents**

   Table 2 shows a small cross section of a database of calculated molecular masses for over 9 primer sets and approximately 30 organisms. The primer sets were derived from rRNA alignment. Examples of regions from rRNA consensus alignments are shown in Figures 1A-1C. Lines with arrows are examples of regions to which intelligent primer pairs for PCR are

30 designed. The primer pairs are >95% conserved in the bacterial sequence database (currently over 10,000 organisms). The intervening regions are variable in length and/or composition, thus providing the base composition "signature" (BCS) for each organism. Primer pairs were chosen so the total length of the amplified region is less than about 80-90 nucleotides. The label for each

primer pair represents the starting and ending base number of the amplified region on the consensus diagram.

Included in the short bacterial database cross-section in Table 2 are many well known pathogens/biowarfare agents (shown in bold/red typeface) such as *Bacillus anthracis* or *Yersinia pestis* as well as some of the bacterial organisms found commonly in the natural environment such as *Streptomyces*. Even closely related organisms can be distinguished from each other by the appropriate choice of primers. For instance, two low G+C organisms, *Bacillus anthracis* and *Staph aureus*, can be distinguished from each other by using the primer pair defined by 16S_1337 or 23S_855 (ΔM of 4 Da).

Table 2: Cross Section Of A Database Of Calculated Molecular Masses[1]

Primer Regions ---> Bug Name	16S_971	16S_1100	16S_1337	16S_1294	16S_1228	23S_1021	23S_855	23S_193	23S_115
Acinetobacter calcoaceticus	55819.1	55004	28448.7	35854.9	51295.4	30299	42354	39557.5	54999
Bacillus anthracis	55005	54388	28448	35238	51296	30295	42651	39560	56850
Bacillus cereus	55622.1	54387.9	28447.8	35854.9	51296.4	30296	42651	39560.5	56850.3
Bordetella bronchiseptica	56857.3	51300.4	28447.9	35857.9	51307.4	30299	42653	39559.5	51920.6
Borrelia burgdorferi	56231.2	55821.1	28440.7	35852.9	51295.4	30297	42029.9	38941.4	52524.8
Brucella abortus	58098	55011	28448	35854	50683				
Campylobacter jejuni	58088.5	54386.9	29061.8	35856.9	50674.3	30294	42032.9	39558.5	45732.5
Chlamydia pneumoniae	55000	55007	29063	35855	50676	30296	42036	38941	56230
Clostridium botulinum	55006	53757	28448	35855	51291	30300	42656	39602	54999
Clostridium difficile	55855.3	54386.9	28444.7	35853.9	51296.4	30294	41417.8	39556.5	55612.2
Enterococcus faecalis	55620.1	54387.9	28447.6	35858.9	51296.4	30297	42652	39559.5	56849.3
Escherichia coli	55622	55009	28445	35857	51301	30301	42656	39562	54999
Francisella tularensis	53769	54386	28445	35856	51298				
Haemophilus influenzae	55620.1	55006	28444.7	35855.9	51298.4	30298	42655	39550.5	55813.1
Klebsiella pneumoniae	55622.1	55008	28442.7	35856.9	51297.4	30300	42655	39562.5	55000
Legionella pneumophila	55618	55626	28446	35857	51303				
Mycobacterium avium	54390.9	55631.1	29064.8	35858.9	51915.5	30298	42658	38942.4	58241.2
Mycobacterium leprae	54389.9	55629.1	29064.8	35860.9	51917.5	30298	42655	39559.5	58240.2
Mycobacterium tuberculosis	54390.9	55629.1	29064.8	35860.9	51301.4	30299	42658	39560.5	58243.2
Mycoplasma genitalium	53143.7	45115.4	29061.8	35854.9	50671.3	30294	43264.1	39558.5	56842.4
Mycoplasma pneumoniae	53143.7	45118.4	29061.8	35854.9	50673.3	30294	43264.1	39559.5	56843.4
Neisseria gonorrhoeae	55627.1	54389.9	28445.7	35855.9	51302.4	30300	42549	39561.5	55000
Pseudomonas aeruginosa	55623	55010	28443	35858	51301	30298	43272	39558	55619
Rickettsia prowazekii	58093	55621	28448	35853	50677	30293	42690	39559	53139
Rickettsia rickettsii	58094	55623	28448	35853	50679	30293	42648	39589	53758
Salmonella typhimurium	55622	55005	28445	35857	51301	30301	42658		
Shigella dysenteriae	55623	55009	28444	35857	51301				
Staphylococcus aureus	56854.3	54386.9	28443.7	35852.9	51294.4	30298	42655	39550.5	57466.4
Streptomyces	54389.9	59341.8	29063.8	35858.9	51300.4			39583.5	55854.3
Treponema pallidum	56245.2	55631.1	28445.7	35851.9	51297.4	30299	42034.9	38939.4	57473.4
Vibrio cholerae	55626	55626	28443	35857	52536	29063	30303	35241	50675
Vibrio parahaemolyticus	54384.9	55626.1	28444.7	34620.7	50064.2				
Yersinia pestis	55620	55626	28443	35857	51299				

[1]Molecular mass distribution of PCR amplified regions for a selection of organisms (rows) across various primer pairs (columns). Pathogens are shown in **bold**. Empty cells indicate presently incomplete or missing data.

Figure 6 shows the use of ESI-FT-ICR MS for measurement of exact mass. The spectra from 46mer PCR products originating at position 1337 of the 16S rRNA from *S. aureus* (upper) and *B. anthracis* (lower) are shown. These data are from the region of the spectrum containing signals from the $[M-8H+]^{8-}$ charge states of the respective 5'-3' strands. The two strands differ by two (AT→CG) substitutions, and have measured masses of 14206.396 and 14208.373 +

0.010 Da, respectively. The possible base compositions derived from the masses of the forward and reverse strands for the *B. anthracis* products are listed in Table 3.

Table 3: Possible base composition for *B. anthracis* products

Calc. Mass	Error	Base Comp.
14208.2935	0.079520	A1 G17 C10 T18
14208.3160	0.056980	A1 G20 C15 T10
14208.3386	0.034440	A1 G23 C20 T2
14208.3074	0.065560	A6 G11 C3 T26
14208.3300	0.043020	A6 G14 C8 T18
14208.3525	0.020480	A6 G17 C13 T10
14208.3751	0.002060	A6 G20 C18 T2
14208.3439	0.029060	A11 G8 C1 T26
14208.3665	0.006520	A11 G11 C6 T18
14208.3890	**0.016020**	**A11 G14 C11 T10**
14208.4116	0.038560	A11 G17 C16 T2
14208.4030	0.029980	A16 G8 C4 T18
14208.4255	0.052520	A16 G11 C9 T10
14208.4481	0.075060	A16 G14 C14 T2
14208.4395	0.066480	A21 G5 C2 T18
14208.4620	0.089020	A21 G8 C7 T10
14079.2624	0.080600	A0 G14 C13 T19
14079.2849	0.058060	A0 G17 C18 T11
14079.3075	0.035520	A0 G20 C23 T3
14079.2538	0.089180	A5 G5 C1 T35
14079.2764	0.066640	A5 G8 C6 T27
14079.2989	0.044100	A5 G11 C11 T19
14079.3214	0.021560	A5 G14 C16 T11
14079.3440	0.000980	A5 G17 C21 T3
14079.3129	0.030140	A10 G5 C4 T27
14079.3354	0.007600	A10 G8 C9 T19
14079.3579	**0.014940**	**A10 G11 C14 T11**
14079.3805	0.037480	A10 G14 C19 T3
14079.3494	0.006360	A15 G2 C2 T27

14079.3719	0.028900	A15 G5 C7 T19
14079.3944	0.051440	A15 G8 C12 T11
14079.4170	0.073980	A15 G11 C17 T3
14079.4084	0.065400	A20 G2 C5 T19
14079.4309	0.087940	A20 G5 C10 T13

Among the 16 compositions for the forward strand and the 18 compositions for the reverse strand that were calculated, only one pair (shown in **bold**) are complementary, corresponding to the actual base compositions of the *B. anthracis* PCR products.

Example 4: BCS of Region from *Bacillus anthracis* and *Bacillus cereus*

A conserved Bacillus region from *B. anthracis* ($A_{14}G_9C_{14}T_9$) and *B. cereus* ($A_{15}G_9C_{13}T_9$) having a C to A base change was synthesized and subjected to ESI-TOF MS. The results are shown in Figure 7 in which the two regions are clearly distinguished using the method of the present invention (MW=14072.26 vs. 14096.29).

Example 5: Identification of additional bioagents

In other examples of the present invention, the pathogen *Vibrio cholera* can be distinguished from *Vibrio parahemolyticus* with $\Delta M > 600$ Da using one of three 16S primer sets shown in Table 2 (16S_971, 16S_1228 or 16S_1294) as shown in Table 4. The two mycoplasma species in the list (*M. genitalium* and *M. pneumoniae*) can also be distinguished from each other, as can the three mycobacteriae. While the direct mass measurements of amplified products can identify and distinguish a large number of organisms, measurement of the base composition signature provides dramatically enhanced resolving power for closely related organisms. In cases such as *Bacillus anthracis* and *Bacillus cereus* that are virtually indistinguishable from each other based solely on mass differences, compositional analysis or fragmentation patterns are used to resolve the differences. The single base difference between the two organisms yields different fragmentation patterns, and despite the presence of the ambiguous/unidentified base N at position 20 in *B. anthracis*, the two organisms can be identified.

Tables 4a-b show examples of primer pairs from Table 1 which distinguish pathogens from background.

Table 4a

Organism name	23S_855	16S_1337	23S_1021
Bacillus anthracis	42650.98	28447.65	30294.98
Staphylococcus aureus	42654.97	28443.67	30297.96

Table 4b

Organism name	16S_971	16S_1294	16S_1228
Vibrio cholerae	55625.09	35856.87	52535.59
Vibrio parahaemolyticus	54384.91	34620.67	50064.19

Table 5 shows the expected molecular weight and base composition of region 16S_1100-1188 in *Mycobacterium avium* and *Streptomyces sp.*

Table 5

Region	Organism name	Length	Molecular weight	Base comp.
16S_1100-1188	*Mycobacterium avium*	82	25624.1728	$A_{16}G_{32}C_{18}T_{16}$
16S_1100-1188	*Streptomyces sp.*	96	29904.871	$A_{17}G_{38}C_{27}T_{14}$

Table 6 shows base composition (single strand) results for 16S_1100-1188 primer amplification reactions different species of bacteria. Species which are repeated in the table (e.g., *Clostridium botulinum*) are different strains which have different base compositions in the 16S_1100-1188 region.

Table 6

Organism name	Base comp.	Organism name	Base comp.
Mycobacterium avium	$A_{16}G_{32}C_{18}T_{16}$	*Vibrio cholerae*	$A_{23}G_{30}C_{21}T_{16}$
Streptomyces sp.	$A_{17}G_{38}C_{27}T_{14}$	*Aeromonas hydrophila*	$A_{23}G_{31}C_{21}T_{15}$
Ureaplasma urealyticum	$A_{18}G_{30}C_{17}T_{17}$	*Aeromonas salmonicida*	$A_{23}G_{31}C_{21}T_{15}$
Streptomyces sp.	$A_{19}G_{36}C_{24}T_{18}$	*Mycoplasma genitalium*	$A_{24}G_{19}C_{12}T_{18}$
Mycobacterium leprae	$A_{20}G_{32}C_{22}T_{16}$	*Clostridium botulinum*	$A_{24}G_{25}C_{18}T_{20}$
M. tuberculosis	$\mathbf{A_{20}G_{33}C_{21}T_{16}}$	*Bordetella bronchiseptica*	$A_{24}G_{26}C_{19}T_{14}$
Nocardia asteroides	$\mathbf{A_{20}G_{33}C_{21}T_{16}}$	*Francisella tularensis*	$A_{24}G_{26}C_{19}T_{19}$
Fusobacterium necroforum	$A_{21}G_{26}C_{22}T_{18}$	**Bacillus anthracis**	$A_{24}G_{26}C_{20}T_{18}$
Listeria monocytogenes	$A_{21}G_{27}C_{19}T_{19}$	**Campylobacter jejuni**	$A_{24}G_{26}C_{20}T_{18}$

-29-

Clostridium botulinum	$A_{21}G_{27}C_{19}T_{21}$	**Staphylococcus aureus**	$A_{24}G_{26}C_{20}T_{18}$
Neisseria gonorrhoeae	$A_{21}G_{28}C_{21}T_{18}$	Helicobacter pylori	$A_{24}G_{26}C_{20}T_{19}$
Bartonella quintana	$A_{21}G_{30}C_{22}T_{16}$	Helicobacter pylori	$A_{24}G_{26}C_{21}T_{18}$
Enterococcus faecalis	$A_{22}G_{27}C_{20}T_{19}$	Moraxella catarrhalis	$A_{24}G_{26}C_{23}T_{16}$
Bacillus megaterium	$A_{22}G_{28}C_{20}T_{18}$	Haemophilus influenzae Rd	$A_{24}G_{28}C_{20}T_{17}$
Bacillus subtilis	$A_{22}G_{28}C_{21}T_{17}$	**Chlamydia trachomatis**	$A_{24}G_{28}C_{21}T_{16}$
Pseudomonas aeruginosa	$A_{22}G_{29}C_{23}T_{15}$	**Chlamydophila pneumoniae**	$A_{24}G_{28}C_{21}T_{16}$
Legionella pneumophila	$A_{22}G_{32}C_{20}T_{16}$	**C. pneumonia AR39**	$A_{24}G_{28}C_{21}T_{16}$
Mycoplasma pneumoniae	$A_{23}G_{20}C_{14}T_{16}$	Pseudomonas putida	$A_{24}G_{29}C_{21}T_{16}$
Clostridium botulinum	$A_{23}G_{26}C_{20}T_{19}$	**Proteus vulgaris**	$A_{24}G_{30}C_{21}T_{15}$
Enterococcus faecium	$A_{23}G_{26}C_{21}T_{18}$	**Yersinia pestis**	$A_{24}G_{30}C_{21}T_{15}$
Acinetobacter calcoaceti	$A_{23}G_{26}C_{21}T_{19}$	**Yersinia pseudotuberculos**	$A_{24}G_{30}C_{21}T_{15}$
Leptospira borgpeterseni	$\mathbf{A_{23}G_{26}C_{24}T_{15}}$	Clostridium botulinum	$A_{25}G_{24}C_{18}T_{21}$
Leptospira interrogans	$\mathbf{A_{23}G_{26}C_{24}T_{15}}$	Clostridium tetani	$A_{25}G_{25}C_{18}T_{20}$
Clostridium perfringens	$A_{23}G_{27}C_{19}T_{19}$	Francisella tularensis	$A_{25}G_{25}C_{19}T_{19}$
Bacillus anthracis	$\mathbf{A_{23}G_{27}C_{20}T_{18}}$	Acinetobacter calcoacetic	$A_{25}G_{26}C_{20}T_{19}$
Bacillus cereus	$\mathbf{A_{23}G_{27}C_{20}T_{18}}$	Bacteriodes fragilis	$A_{25}G_{27}C_{16}T_{22}$
Bacillus thuringiensis	$\mathbf{A_{23}G_{27}C_{20}T_{18}}$	Chlamydophila psittaci	$A_{25}G_{27}C_{21}T_{16}$
Aeromonas hydrophila	$A_{23}G_{29}C_{21}T_{16}$	Borrelia burgdorferi	$A_{25}G_{29}C_{17}T_{19}$
Escherichia coli	$A_{23}G_{29}C_{21}T_{16}$	Streptobacillus monilifor	$A_{26}G_{26}C_{20}T_{16}$
Pseudomonas putida	$A_{23}G_{29}C_{21}T_{17}$	Rickettsia prowazekii	$A_{26}G_{28}C_{18}T_{18}$
Escherichia coli	$\mathbf{A_{23}G_{29}C_{22}T_{15}}$	Rickettsia rickettsii	$A_{26}G_{28}C_{20}T_{16}$
Shigella dysenteriae	$\mathbf{A_{23}G_{29}C_{22}T_{15}}$	Mycoplasma mycoides	$A_{28}G_{23}C_{16}T_{20}$

The same organism having different base compositions are different strains. Groups of organisms which are highlighted or in italics have the same base compositions in the amplified region. Some of these organisms can be distinguished using multiple primers. For example,
5 *Bacillus anthracis* can be distinguished from *Bacillus cereus* and *Bacillus thuringiensis* using the primer 16S_971-1062 (Table 7). Other primer pairs which produce unique base composition signatures are shown in Table 6 (bold). Clusters containing very similar threat and ubiquitous non-threat organisms (e.g. *anthracis* cluster) are distinguished at high resolution with focused sets of primer pairs. The known biowarfare agents in Table 6 are *Bacillus anthracis, Yersinia*
10 *pestis, Francisella tularensis* and *Rickettsia prowazekii.*

Table 7

Organism	16S_971-1062	16S_1228-1310	16S_1100-1188
Aeromonas hydrophila	$A_{21}G_{29}C_{22}T_{20}$	$A_{22}G_{27}C_{21}T_{13}$	$A_{23}G_{31}C_{21}T_{15}$
Aeromonas salmonicida	$A_{21}G_{29}C_{22}T_{20}$	$A_{22}G_{27}C_{21}T_{13}$	$A_{23}G_{31}C_{21}T_{15}$
Bacillus anthracis	$\mathbf{A_{21}G_{27}C_{22}T_{22}}$	$A_{24}G_{22}C_{19}T_{18}$	$A_{23}G_{27}C_{20}T_{18}$
Bacillus cereus	$A_{22}G_{27}C_{21}T_{22}$	$A_{24}G_{22}C_{19}T_{18}$	$A_{23}G_{27}C_{20}T_{18}$
Bacillus thuringiensis	$A_{22}G_{27}C_{21}T_{22}$	$A_{24}G_{22}C_{19}T_{18}$	$A_{23}G_{27}C_{20}T_{18}$
Chlamydia trachomatis	$\mathbf{A_{22}G_{26}C_{20}T_{23}}$	$A_{24}G_{23}C_{19}T_{16}$	$A_{24}G_{28}C_{21}T_{16}$
Chlamydia pneumoniae AR39	$A_{26}G_{23}C_{20}T_{22}$	$A_{26}G_{22}C_{16}T_{18}$	$A_{24}G_{28}C_{21}T_{16}$
Leptospira borgpetersenii	$A_{22}G_{26}C_{20}T_{21}$	$A_{22}G_{25}C_{21}T_{15}$	$A_{23}G_{26}C_{24}T_{15}$
Leptospira interrogans	$A_{22}G_{26}C_{20}T_{21}$	$A_{22}G_{25}C_{21}T_{15}$	$A_{23}G_{26}C_{24}T_{15}$
Mycoplasma genitalium	$A_{28}G_{23}C_{15}T_{22}$	$\mathbf{A_{30}G_{18}C_{15}T_{19}}$	$\mathbf{A_{24}G_{19}C_{12}T_{18}}$
Mycoplasma pneumoniae	$A_{28}G_{23}C_{15}T_{22}$	$\mathbf{A_{27}G_{19}C_{16}T_{20}}$	$\mathbf{A_{23}G_{20}C_{14}T_{16}}$
Escherichia coli	$\mathbf{A_{22}G_{28}C_{20}T_{22}}$	$A_{24}G_{25}C_{21}T_{13}$	$A_{23}G_{29}C_{22}T_{15}$
Shigella dysenteriae	$\mathbf{A_{22}G_{28}C_{21}T_{21}}$	$A_{24}G_{25}C_{21}T_{13}$	$A_{23}G_{29}C_{22}T_{15}$
Proteus vulgaris	$\mathbf{A_{23}G_{26}C_{22}T_{21}}$	$A_{26}G_{24}C_{19}T_{14}$	$A_{24}G_{30}C_{21}T_{15}$
Yersinia pestis	$A_{24}G_{25}C_{21}T_{22}$	$A_{25}G_{24}C_{20}T_{14}$	$A_{24}G_{30}C_{21}T_{15}$
Yersinia pseudotuberculosis	$A_{24}G_{25}C_{21}T_{22}$	$A_{25}G_{24}C_{20}T_{14}$	$A_{24}G_{30}C_{21}T_{15}$
Francisella tularensis	$\mathbf{A_{20}G_{25}C_{21}T_{23}}$	$A_{23}G_{26}C_{17}T_{17}$	$\mathbf{A_{24}G_{26}C_{19}T_{19}}$
Rickettsia prowazekii	$\mathbf{A_{21}G_{26}C_{24}T_{25}}$	$A_{24}G_{23}C_{16}T_{19}$	$\mathbf{A_{26}G_{28}C_{18}T_{18}}$
Rickettsia rickettsii	$\mathbf{A_{21}G_{26}C_{25}T_{24}}$	$A_{24}G_{24}C_{17}T_{17}$	$\mathbf{A_{26}G_{28}C_{20}T_{16}}$

The sequence of B. anthracis and B. cereus in region 16S_971 is shown below. Shown in bold is the single base difference between the two species which can be detected using the methods of the present invention. B. anthracis has an ambiguous base at position 20.

B.anthracis_16S_971

GCGAAGAACCUUACCAGGUMUUGACAUCCUCUGA**C**AACCCUAGAGAUAGGGCUUC
UCCUUCGGGAGCAGAGUGACAGGUGGUGCAUGGUU (SEQ ID NO:1)

B.cereus_16S_971

GCGAAGAACCUUACCAGGUCUUGACAUCCUCUGAAAACCCUAGAGAUAGGGCUUC
UCCUUCGGGAGCAGAGUGACAGGUGGUGCAUGGUU (SEQ ID NO:2)

Example 6: ESI-TOF MS of sspE 56-mer Plus Calibrant

The mass measurement accuracy that can be obtained using an internal mass standard in the ESI-MS study of PCR products is shown in Fig.8. The mass standard was a 20-mer phosphorothioate oligonucleotide added to a solution containing a 56-mer PCR product from the
5 *B. anthracis* spore coat protein sspE. The mass of the expected PCR product distinguishes *B. anthracis* from other species of Bacillus such as *B. thuringiensis* and *B. cereus*.

Example 7: *B. anthracis* ESI-TOF Synthetic 16S_1228 Duplex

An ESI-TOF MS spectrum was obtained from an aqueous solution containing 5 µM
10 each of synthetic analogs of the expected forward and reverse PCR products from the nucleotide 1228 region of the *B. anthracis* 16S rRNA gene. The results (Fig. 9) show that the molecular weights of the forward and reverse strands can be accurately determined and easily distinguish the two strands. The $[M-21H^+]^{21-}$ and $[M-20H^+]^{20-}$ charge states are shown.

15 **Example 8: ESI-FTICR-MS of Synthetic *B. anthracis* 16S_1337 46 Base Pair Duplex**

An ESI-FTICR-MS spectrum was obtained from an aqueous solution containing 5 µM each of synthetic analogs of the expected forward and reverse PCR products from the nucleotide 1337 region of the *B. anthracis* 16S rRNA gene. The results (Fig. 10) show that the molecular weights of the strands can be distinguished by this method. The $[M-16H^+]^{16-}$ through $[M-
20 10H^+]^{10-}$ charge states are shown. The insert highlights the resolution that can be realized on the FTICR-MS instrument, which allows the charge state of the ion to be determined from the mass difference between peaks differing by a single 13C substitution.

**Example 9: ESI-TOF MS of 56-mer Oligonucleotide from saspB Gene of *B. anthracis* with
25 Internal Mass Standard**

ESI-TOF MS spectra were obtained on a synthetic 56-mer oligonucleotide (5 µM) from the saspB gene of *B. anthracis* containing an internal mass standard at an ESI of 1.7 µL/min as a function of sample consumption. The results (Fig. 11) show that the signal to noise is improved as more scans are summed, and that the standard and the product are visible after only 100 scans.
30

Example 10: ESI-TOF MS of an Internal Standard with Tributylammonium (TBA)-trifluoroacetate (TFA) Buffer

An ESI-TOF-MS spectrum of a 20-mer phosphorothioate mass standard was obtained following addition of 5 mM TBA-TFA buffer to the solution. This buffer strips charge from the oligonucleotide and shifts the most abundant charge state from $[M-8H^+]^{8-}$ to $[M-3H^+]^{3-}$ (Fig. 12).

Example 11: Master Database Comparison

The molecular masses obtained through Examples 1-10 are compared to molecular masses of known bioagents stored in a master database to obtain a high probability matching molecular mass.

Example 12: Master Data Base Interrogation over the Internet

The same procedure as in Example 11 is followed except that the local computer did not store the Master database. The Master database is interrogated over an internet connection, searching for a molecular mass match.

Example 13: Master Database Updating

The same procedure as in example 11 is followed except the local computer is connected to the internet and has the ability to store a master database locally. The local computer system periodically, or at the user's discretion, interrogates the Master database, synchronizing the local master database with the global Master database. This provides the current molecular mass information to both the local database as well as to the global Master database. This further provides more of a globalized knowledge base.

Example 14: Global Database Updating

The same procedure as in example 13 is followed except there are numerous such local stations throughout the world. The synchronization of each database adds to the diversity of information and diversity of the molecular masses of known bioagents.

Example 15: Environmental Sampling Protocol for Anthrax Spores Using Non-Cotton Swabs

The use of non-cotton, sterile swabs to collect environmental samples is the preferred method because it reduces the amount of normal background contamination and improves the recovery of anthrax spores (see, for example, www.dhmh.state.md.us/labs/html/Terrorism/Biological/BT_env_samp_prtcl.html).

The following protocol is used to collect samples from small non-porous surfaces or objects for culture and identification of *Bacillus anthracis*. Non-powdered gloves, a disposable gown, and facemask are worn during the collection of swab samples from the environment. A sterile Dacron or Rayon dry swab (non-cotton swab) is used to collect environmental samples.

5 Alternatively, a swab moistened with sterile saline or distilled water is used to collect samples from computer keyboards, desks, mail sorting areas, and other sites within a building or work facility. A separate swab is used for each site. In an open area (desks, work table, mail sorting area, etc.) a swab is taken of a 10 x 10 inch square area per swab. Each swab is placed in a 15 ml sterile tube, the shaft of swab is snapped off at the lip of the tube, and the tube is closed. Each

10 tube is labeled appropriately and placed in a self-sealable plastic bag. Multiple tubes can be placed in the same bag. Each tube is labeled with the date, facility location, and sampling site. The outside of the sealed bag is cleaned by wiping with 10% bleach solution. Place clean, sealed bags into an unused similar self-sealing bag before delivery. Appropriate chain-of-custody documentation and procedure is constantly maintained.

15

Example 16: Environmental Sampling Protocol for Microbial Soil Samples

Soil cores are collected to a depth of 8 cm, using an 8 cm diameter soil corer. Soils for analysis of root biomass and composition are collected separately. A single 3.5 cm diameter core is collected to a depth of 16 cm. Three 3.5 cm cores are also taken. These cores are divided

20 into 0-1, 1-8, and 8-16 cm depth increments. Polyethylene bags containing soil cores are handled to minimize disturbance and changes in soil moisture and temperature. While sampling, any unusual features of the area where the core was taken are noted.

Microbial Group soil cores are sieved (4 mm) and the plant material discarded.

25 **Example 17: Estimation of the Sensitivity of Detection of Spores in Water Samples**

In order to determine the detection limits of the system against organisms whose genetic material is difficult to extract, dilution to extinction experiments were performed using *Bacillis thuringiensis israeliensis* (Bti) spores. Highly purified spores were added to water at concentrations that would give between 0 and $1x10^5$ spores per PCR reaction. Samples were

30 subjected to a lysis protocol. Genomic material was isolated on a Qiagen 8000 Biorobot using a modified robotic version of the DNeasy protocol from Qiagen. The DNeasy protocol was optimized for bacterial lysis with the addition of lysozyme and was modified for use on the BioRobot 8000. PCR reactions were assembled using intelligent primers for bacteria.

These reactions were set up to run automatically using a Perkin Elmer MPII robot with an additional gripper arm. Automated genome isolation, reaction setup, and subsequent PCR of the samples results in the production of uniform, reproducible amounts of amplicon material. PCR products were desalted and purified using a recently published procedure (Jiang, Y. and

5   S.A. Hofstadler, *Anal. Biochem.*, **2002** in press) and were analyzed by electrospray ionization (ESI)-FTICR mass spectrometry. Figure 18 shows the spectra resulting from a single primer pair at each concentration of Bti spores used. Using a maximum likelihood processor, we reproducibly detected Bti genomic material in samples containing as few as 100 spores. This processor readily detects signals that are not easily discernable relative to the noise by eye.

10 Similar results were obtained from the processing of *Bacillus anthracis* (Sterne strain) spores (data not shown).

Example 18: Identification of Bioagents in air samples.

Multi-hour air samples from indoor or outdoor air were obtained using a Spin-Con

15 sampler which can sample up to 18×10^6 L of air. The air samples were processed as 10 ml samples and spiked with *Bacillus anthracis* spores or *Bacillus thuringiensis israeliensis* (Bti) spores. The samples were filtered and lysed via bead-beating. Nucleic acids were isolated from the samples by standard methods and amplified by PCR using intelligent primers. Samples were then desalted and analyzed by FT-ICR mass spectrometry according to methods outlined in

20 example 2. Analysis of the molecular masses enabled detection and identification of the *Bacillis anthracis* spores or *Bacillis thuringiensis israeliensis* (Bti) spores. Figure 19 is an ROC curve that indicates that a spore spiked sample (Pd sample threat) is easily distinguished from background bioagents (Pfa sample threat and Pfa system). The data set indicates that the present method of air sample detection and identification of bioagents provides a reliable means of air
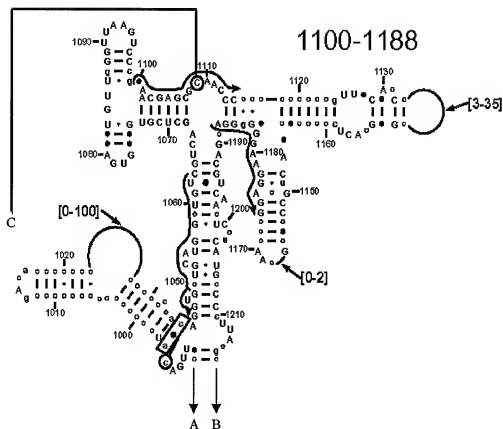
25 sample surveillance.

Various modifications of the invention, in addition to those described herein, will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. Each reference cited in the present application is incorporated herein by reference in its entirety. The following US applications are

30 incorporated herein by reference in their entirety: Serial No. 10/660,996 filed September 12, 2003, Serial No. 10/326,047 filed December 18, 2002, Serial No. 10/323,186 filed December 18, 2002, Serial No. 10/326,642 filed December 18, 2002, Serial No. 60/431,319 filed December 6, 2002, Serial No. 60/443,788 filed January 30, 2003, Serial No. 60/453,607 filed March 10, 2003, and Serial No. 60/454,165 filed March 12, 2003.

What is claimed is:

1.   A method of identifying a bioagent in a sample comprising the steps of:
        determining a first molecular mass of a first amplification product of a first bioagent identifying amplicon from the sample and comparing the first molecular mass to a second
5   molecular mass of a second bioagent identifying amplicon, wherein both first and second bioagent identifying amplicons are correlative.

2.   A method of claim 1 wherein the sample is an environmental sample.

3.   A method of claim 2 wherein the environmental sample is an air sample.

4.   A method of claim 2 wherein the environmental sample is a water sample.

10 5.   A method of claim 2 wherein the environmental sample is a soil sample.

6.   A method of claim 2 wherein the environmental sample is a surface swab sample.

7.   A method of claim 2 wherein the environmental sample is from a building or a container.

8.   A method of claim 1 wherein the sample is a product sample.

15 9.   A method of claim 8 wherein the product sample is a foodstuff.

10.   A method of claim 8 wherein the product sample is a cosmetic.

11.   A method of claim 1 wherein the second molecular mass of the second bioagent identifying amplicon is indexed to a defined bioagent and contained in a database.

12.   A method of claim 1 wherein a first base composition signature is determined from the
20 first molecular mass of the first amplification product and wherein the first base composition signature is compared to a second base composition signature determined for the second bioagent identifying amplicon.

13.   A method of claim 12 wherein the second base composition signature is indexed to a defined bioagent and contained in a database.

25 14.   A method of claim 1 wherein the molecular mass of the amplification product is determined by ESI-TOF mass spectrometry.

15.   A method of claim 1 wherein the molecular mass of the amplification product is determined by ESI-FTICR mass spectrometry.

16.   A method of claim 15 wherein a mass spectrum obtained in determination of the
30 molecular mass provides a measure of the quantity of the bioagent in the sample.

17.   A method of claim 1 wherein the bioagent is a bacterium, virus, mold, fungus or parasite.

18.   The method of claim 17 wherein the mold is a household mold.

19.   The method of claim 18 wherein the household mold is *Stachybotrys, Cladosporium, Penicillium,* or *Alternaria.*

20.   A method of monitoring a bioremediation process by identifying bioagents in a sample comprising the steps of:

5   determining a first molecular mass of a first amplification product of a first bioagent identifying amplicon from the sample and comparing the first molecular mass to a second molecular mass of a second bioagent identifying amplicon wherein, both first and second bioagent identifying amplicons are correlative.

21.   A method of claim 20 wherein the second molecular mass of the second bioagent

10 identifying amplicon is indexed to a defined bioagent and contained in a database.

22.   A method of claim 20 wherein a first base composition signature is determined from the first molecular mass of the first amplification product and wherein the first base composition signature is compared to a second base composition signature determined for the second bioagent identifying amplicon.

15 23.   A method of claim 22 wherein the second base composition signature is indexed to a defined bioagent and contained in a database.

24.   A method of claim 20 wherein the molecular mass of the amplification product is determined by ESI-TOF mass spectrometry.

25.   A method of claim 20 wherein the molecular mass of the amplification product is

20 determined by ESI-FTICR mass spectrometry.

26.   A method of claim 25 wherein a mass spectrum obtained in determination of the molecular mass provides a measure of the quantity of the bioagent in the sample.

27.   A method of claim 20 wherein the bioagent is a bacterium or a fungus.
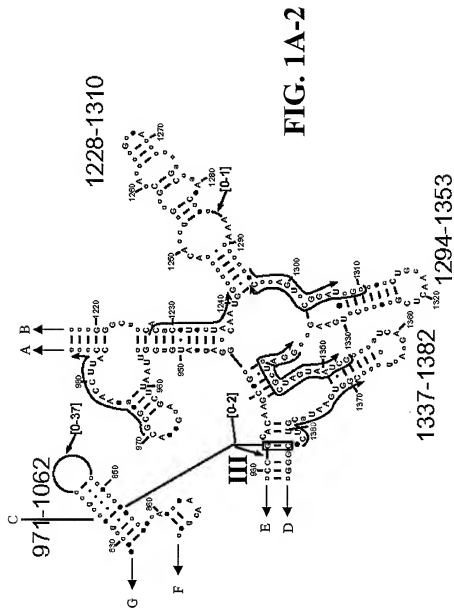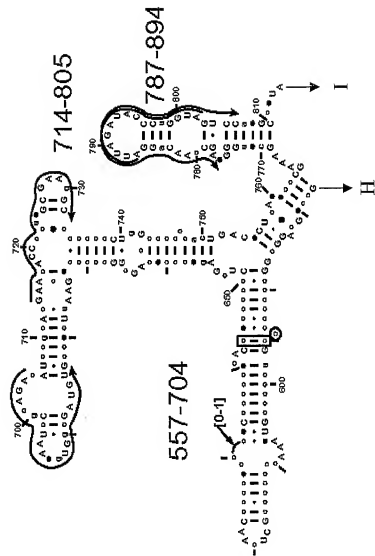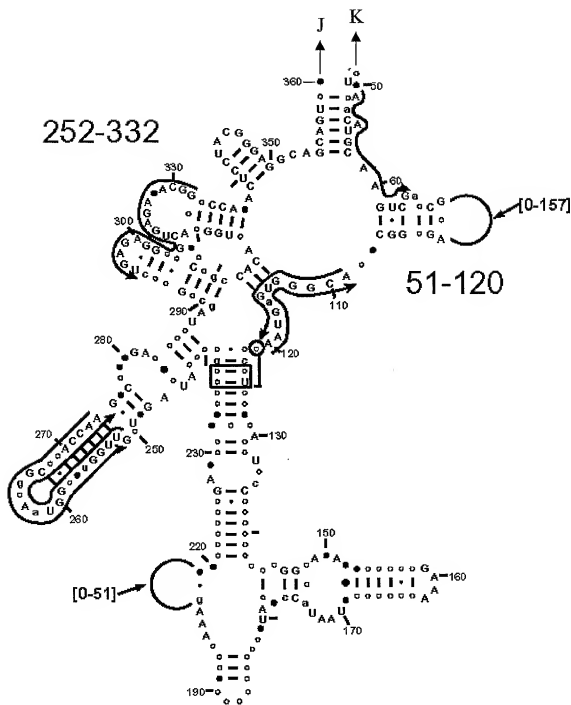
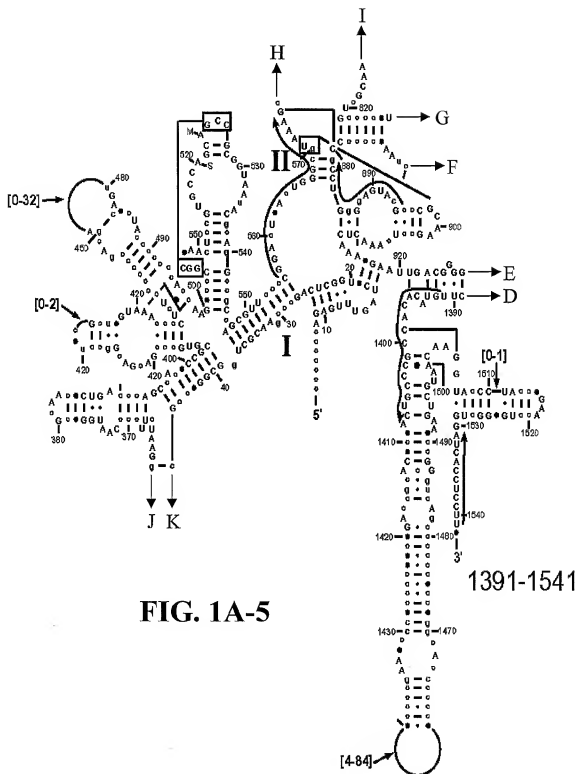# FIG. 1A-1

FIG. 1A-2

FIG. 1A-3

FIG. 1A-4

FIG. 1A-5

FIG. 1B

FIG. 1C

Note: TA in E. coli replaced with CC so that the reverse complement becomes GG to match with predominant YY

1778-1841

1946-2075

[0-4]   [0-6]

[6-53]

IV

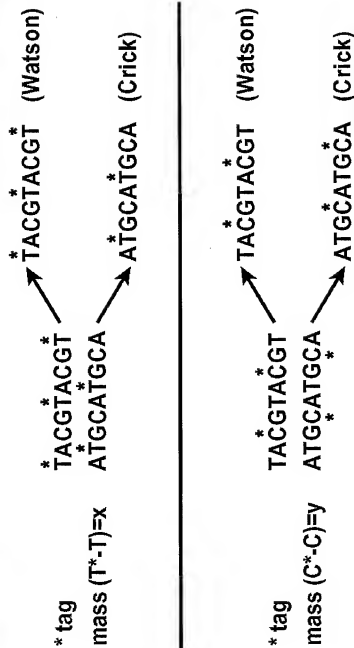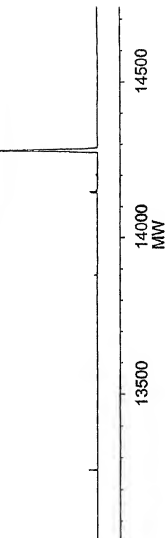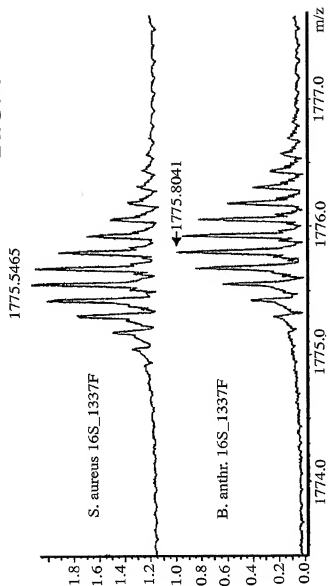FIG. 1D
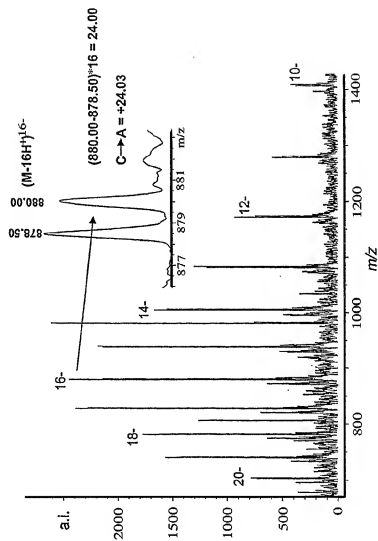
FIG. 1E

FIG. 1F

FIG. 1G

FIG. 1H

FIG. 2
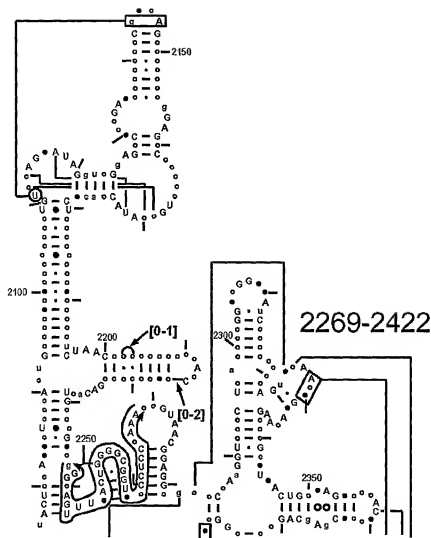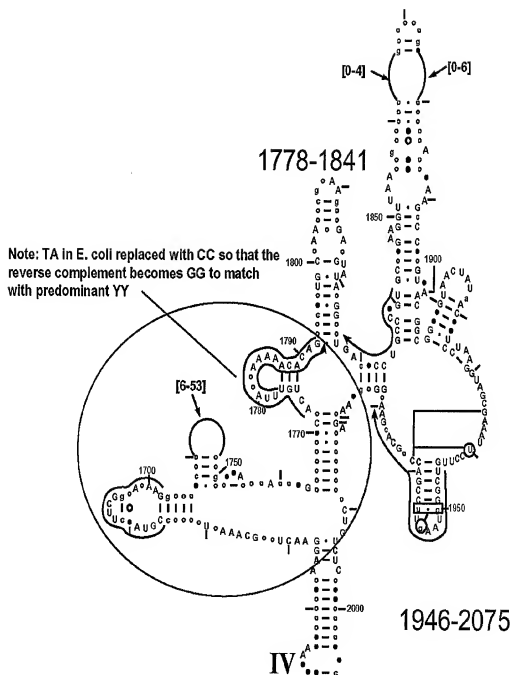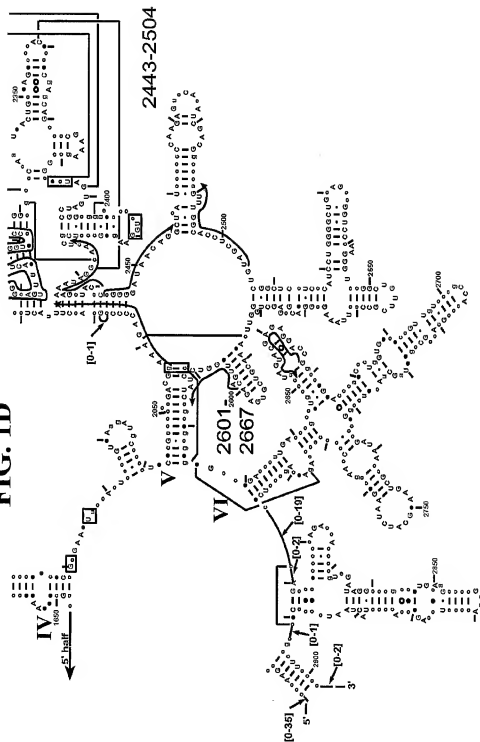
FIG. 3
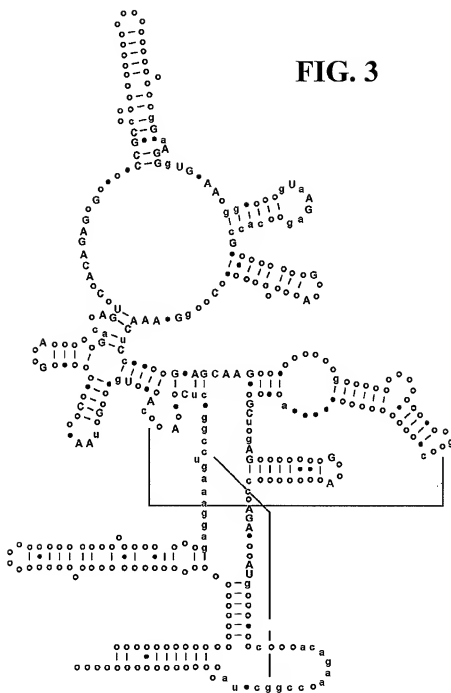
FIG. 4

FIG. 5

B. anthracis $(A_{14}G_9C_{14}T_9)$ MW_{meas} = 14072.2

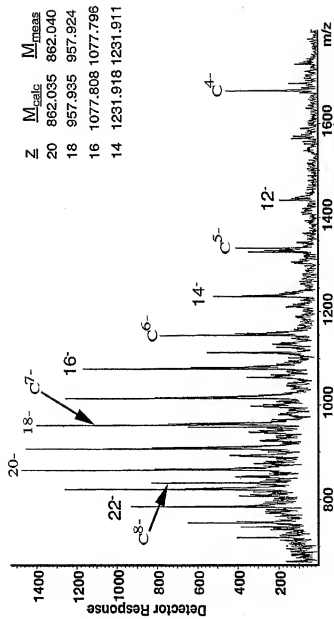B. anthracis* $(A_1A*_{13}G_9C_{14}T_9)$ MW_{meas} = 14280.9

FIG. 6

FIG. 7

FIG. 8

ESI-TOF MS of sspE 56mer + Calibrant

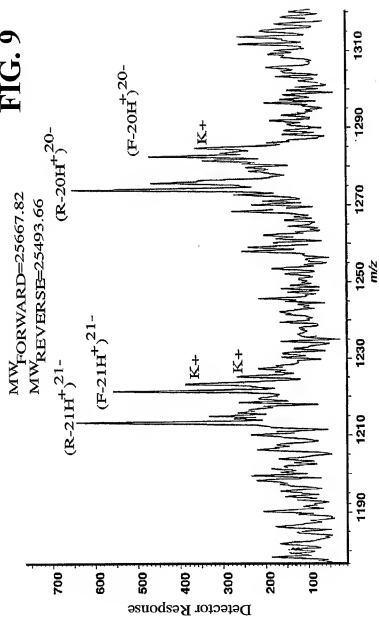| $\underline{z}$	$\underline{M_{calc}}$	$\underline{M_{meas}}$
20	862.035	862.040
18	957.935	957.924
16	1077.808	1077.796
14	1231.918	1231.911

FIG. 9

21/30

FIG. 10

FIG. 11

FIG. 12

FIG. 13
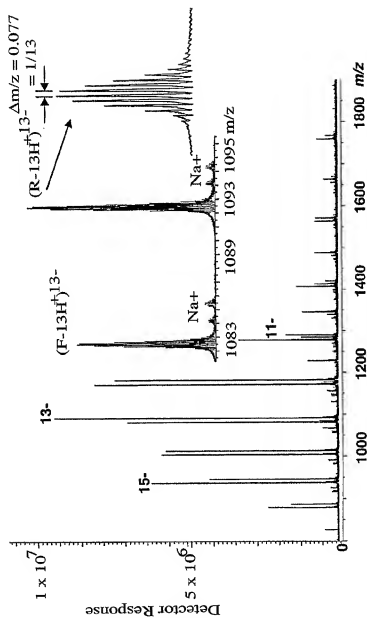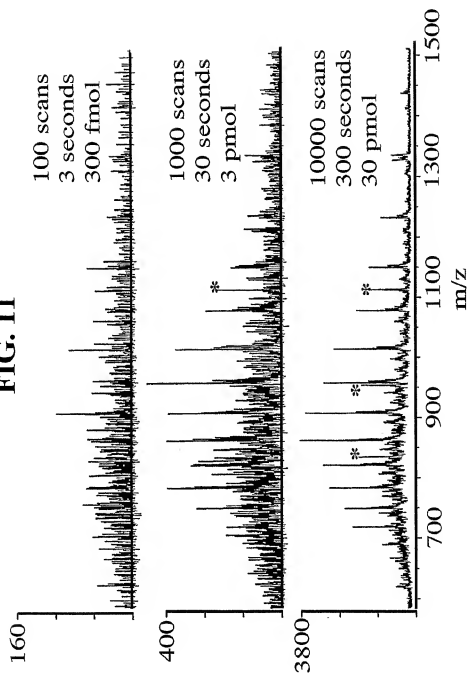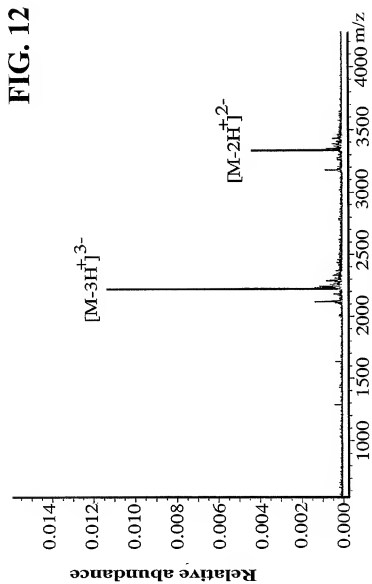
FIG. 14
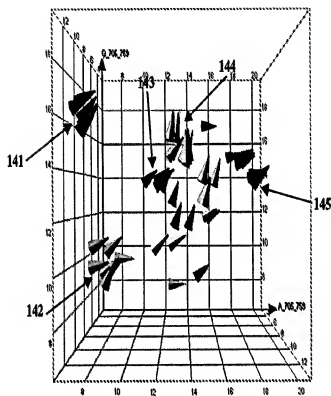
Picorno RdRp 705-759

■ Encephalomyocarditis virus □ Foot-and-mouth disease virus ▨ Polio □ Rhinovirus
□ Enterovirus ■ Hepatitis A virus ■ Porcine enterovirus ▨ Simian Hepatitis A

FIG. 15

Immunodeficiency viruses
● Bovine IV
○ HUMAN IV type 2
○ Human IV type 1
○ Simian IV
○ Feline IV
○ Chimpanzee IV

Hepatitis viruses
▲ Human Hepatitis B
△ Snow Goose Hepatitis B
△ Duck Hepatitis B
△ Woodchuck Hepatitis B
△ Woodchuck Hepatitis 7

FIG. 16

Flavi RdRp 2453-2493

☐ Dengue virus type ☐ Japanese encephalitis virus ☐ Tick-borne encephalitis virus

☐ Dengue virus type ■ Kunjin virus ☐ West Nile virus

■ Dengue virus type ☐ Murray valley encephalitis virus ☐ Yellow fever virus

FIG. 17
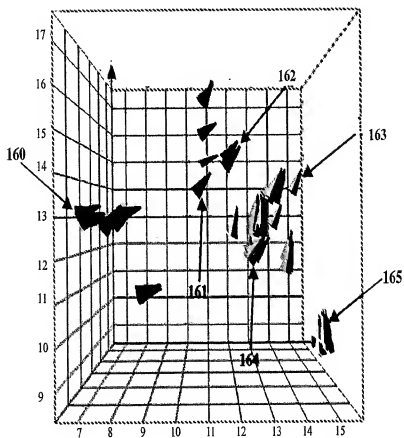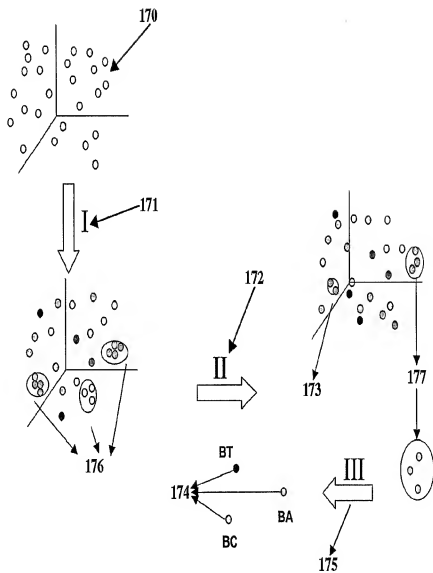
Figure 18

Figure 19
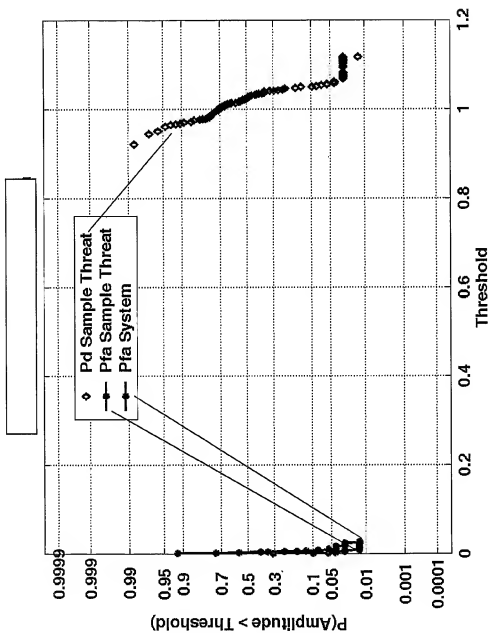
                          SEQUENCE LISTING

<110> Isis Pharmaceuticals, Inc.
      Ecker, David J.
      Griffey, Richard H.
      Sampath, Rangarajan
      Hofstadler, Steven A.
      McNeil, John
      Crooke, Stanley T.
      Blyn, Lawrence B.
      Hall, Thomas A.

<120> Methods For Rapid Detection And Identification Of Bioagents For
      Environmental And Product Testing

<130> IBIS0064-500WO (DIBIS-0002WO.P4)

<150> 10/660,996
<151> 2003-09-12

<150> 60/431,319
<151> 2002-12-06

<150> 10/326,047
<151> 2002-12-18

<150> 10/323,186
<151> 2002-12-18

<150> 10/326,642
<151> 2002-12-18

<150> 60/443,788
<151> 2003-01-30

<150> 60/453,607
<151> 2003-03-10

<150> 60/454,165
<151> 2003-03-12

<160> 4

<170> PatentIn version 3.0

<210> 1
<211> 90
<212> RNA
<213> Bacillus anthracis

<220>
<221> misc_feature
<222> (20)..(20)
<223> N = A, U, G or C

<400> 1
gcgaagaacc uuaccaggun uugacauccu cugacaaccc uagagauagg gcuucuccuu   60

cgggagcaga gugacaggug gugcaugguu   90

                                    1

```
<210> 2
<211> 90
<212> RNA
<213> Bacillus cereus

<400> 2
gcgaagaacc uuaccagguc uugacauccu cugaaaaccc uagagauagg gcuucuccuu      60

cgggagcaga gugacaggug gugcaugguu                                       90

<210> 3
<211> 1542
<212> RNA
<213> Artificial Sequence

<220>
<221> misc_feature
<223> 16S rRNA consensus sequence

<220>
<221> misc_feature
<222> (1)..(7)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (19)..(19)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (28)..(30)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (43)..(45)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (48)..(48)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (50)..(50)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (52)..(52)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (66)..(66)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (69)..(100)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (103)..(103)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (107)..(108)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (121)..(122)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (124)..(124)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (126)..(129)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (131)..(132)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (134)..(134)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (137)..(145)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (148)..(148)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (150)..(150)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (152)..(158)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (163)..(169)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (177)..(178)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (181)..(194)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (199)..(226)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (229)..(237)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (239)..(242)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (245)..(245)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (248)..(248)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (250)..(250)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (257)..(258)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (264)..(264)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (268)..(269)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (276)..(276)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (278)..(280)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (283)..(286)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (291)..(291)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (293)..(294)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (303)..(304)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (306)..(307)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (309)..(309)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (316)..(316)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (320)..(320)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (328)..(328)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (333)..(333)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (337)..(337)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (359)..(360)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (366)..(366)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (369)..(371)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (378)..(379)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (381)..(381)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (384)..(385)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (390)..(392)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (396)..(396)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (398)..(399)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (407)..(409)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (412)..(412)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (415)..(415)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (418)..(419)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (421)..(423)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (425)..(425)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (427)..(427)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (433)..(435)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (438)..(438)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (440)..(446)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (449)..(449)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (452)..(479)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (484)..(485)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (488)..(494)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (496)..(497)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (501)..(503)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (508)..(508)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (513)..(513)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (538)..(538)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (542)..(543)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (546)..(546)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (553)..(555)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (560)..(560)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (562)..(562)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (564)..(564)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (576)..(576)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (578)..(580)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (582)..(582)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (586)..(586)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (589)..(596)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (599)..(603)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (606)..(606)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (610)..(616)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (620)..(620)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (624)..(633)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (635)..(641)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (644)..(650)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (653)..(653)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

<222> (657)..(662)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (665)..(665)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (668)..(673)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (679)..(682)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (689)..(689)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (694)..(694)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (698)..(698)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (701)..(701)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (705)..(705)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (708)..(709)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (711)..(711)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (713)..(713)
<223> N= A, U, G or C

<220>
<221> misc_feature

```
<222> (717)..(717)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (721)..(722)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (724)..(724)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (733)..(738)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (743)..(748)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (755)..(755)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (758)..(758)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (760)..(763)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (771)..(771)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (776)..(776)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (780)..(780)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (808)..(808)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (811)..(812)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (819)..(819)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (822)..(826)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (828)..(831)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (833)..(835)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (837)..(859)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (861)..(863)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (868)..(870)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (874)..(878)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (895)..(896)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (903)..(904)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (906)..(906)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

<222> (916)..(916)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (929)..(929)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (932)..(932)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (941)..(941)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (943)..(943)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (948)..(948)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (955)..(955)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (965)..(965)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (967)..(968)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (974)..(974)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (976)..(976)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (986)..(990)
<223> N= A, U, G or C

<220>
<221> misc_feature

```
<222> (998)..(1012)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1015)..(1015)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1017)..(1043)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1051)..(1051)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1059)..(1059)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1075)..(1076)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1082)..(1082)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1100)..(1100)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1115)..(1123)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1127)..(1127)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1129)..(1129)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1131)..(1131)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1133)..(1141)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1143)..(1143)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1145)..(1145)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1150)..(1156)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1163)..(1165)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1167)..(1168)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1171)..(1173)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1183)..(1183)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1189)..(1189)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1198)..(1198)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1201)..(1201)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1207)..(1207)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1214)..(1214)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1216)..(1219)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1225)..(1225)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1231)..(1231)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1233)..(1233)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1243)..(1247)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1251)..(1252)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1254)..(1254)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1256)..(1257)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1260)..(1260)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1262)..(1265)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1267)..(1268)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1270)..(1274)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1278)..(1278)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1281)..(1281)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1283)..(1286)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1290)..(1294)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1297)..(1298)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1302)..(1302)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1308)..(1308)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1310)..(1313)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1324)..(1327)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1329)..(1329)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1335)..(1336)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1340)..(1340)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1354)..(1356)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1362)..(1362)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1364)..(1364)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1366)..(1368)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1383)..(1383)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1388)..(1388)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1409)..(1411)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1414)..(1414)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1416)..(1417)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1420)..(1428)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1431)..(1432)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1436)..(1447)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1449)..(1454)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1456)..(1465)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1467)..(1467)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1469)..(1469)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1472)..(1481)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1484)..(1484)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1489)..(1491)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1508)..(1508)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1511)..(1511)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1514)..(1516)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1520)..(1521)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1524)..(1524)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1527)..(1527)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1542)..(1542)
<223> N= A, U, G or C

<400> 3
nnnnnnnaga guuugaucnu ggcucagnnn gaacgcuggc ggnnngcnun anacaugcaa      60

gucgancgnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn agnggcnnac gggugaguaa     120

nncnunnnna nnunccnnnn nnnnnnggnan annnnnnnga aannnnnnnu aauaccnnau    180

nnnnnnnnnn nnnnaaagnn nnnnnnnnnn nnnnnnnnnn nnnnnngann nnnnnnngnn     240

nnaunagnun guuggunngg uaanggcnna ccaagncnnn gannnnuagc ngnncugaga     300

ggnngnncng ccacanuggn acugaganac ggnccanacu ccuacgggag gcagcagunn     360

ggaaunuunn ncaaugnnng naanncugan nnagcnannc cgcgugnnng anganggnnu     420

nnngnungua aannncunun nnnnnngang annnnnnnnn nnnnnnnnnn nnnnnnnnnu     480

gacnnuannn nnnnannaag nnncggcnaa cuncgugcca gcagccgcgg uaauacgnag     540

gnngcnagcg uunnncggan unanugggcg uaaagngnnn gnaggnggnn nnnnnngunn     600

nnngunaaaa nnnnnngcun aacnnnnnnn nnncnnnnnn nacnnnnnnn cungagnnnn     660

nnagngnnnn nnngaauunn nnguguagng gugnaauncg naganaunng nangaanacc     720

nnnngcgaag gcnnnnnncu ggnnnnnnac ugacncunan nnncgaaagc nugggnagcn     780

aacaggauua gauacccugg uagccangc nnuaaacgnu gnnnnnunnn ngnnngnnnn      840

nnnnnnnnnn nnnnnnnnna nnnaacgnnn uaannnnncc gccugggag uacgnncgca      900

agnnnunaaac ucaaangaau ugacgggggc cngcacaagc ngnggagnau guggnuuaau    960

ucgangnnac gcgnanaacc uuaccnnnnn uugacaunnn nnnnnnnnnn nnganannnn    1020

nnnnnnnnnn nnnnnnnnnn nnnacagug nugcauggnu gucgucagcu cgugnnguga    1080

gnuguugggu uaaguccgn aacgagcgca acccnnnnnn nnnguucna ncnnnnnnnn     1140

ngngnacucn nnnnnnacug ccnnngnnaa nnnggaggaa ggnggggang acgucaanuc    1200

nucaugnccc uuangnnnng ggcuncacac nuncuacaau ggnnnnnaca nngngnngcn    1260

annnnngnnan nnnnagcnaa ncnnnnnaaan nnnnucnnag uncggaugn nnncugcaac    1320

ucgnnnncnu gaagnnggan ucgcuaguaa ucgnnnauca gnangnnncg gugaauacgu    1380

ucncgggncu uguacacacc gcccgucann ncangnnagn nnnnnnnncc nnaagnnnnn    1440
```

nnnnnnncnn nnnngnnnnn nnnnncnang gnnnnnnnnn nganugggnn naagucguaa   1500

caagguancc nuannngaan nugnggnugg aucaccuccu un   1542

<210> 4
<211> 2904
<212> RNA
<213> Artificial Sequence

<220>
<221> misc_feature
<223> 23S rRNA consensus sequence

<220>
<221> misc_feature
<222> (1)..(4)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (8)..(12)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (16)..(16)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (18)..(22)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (34)..(34)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (38)..(43)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (46)..(46)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (50)..(50)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (57)..(57)
<223> N= A, U, G or C

```
<220>
<221> misc_feature
<222> (59)..(65)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (67)..(68)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (70)..(72)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (74)..(75)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (77)..(79)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (82)..(83)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (86)..(87)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (89)..(96)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (98)..(102)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (104)..(104)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (107)..(109)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (111)..(111)
<223> N= A, U, G or C
```

```
<220>
<221> misc_feature
<222> (113)..(113)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (125)..(125)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (131)..(148)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (150)..(177)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (179)..(181)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (184)..(188)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (192)..(192)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (203)..(203)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (208)..(212)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (218)..(218)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (224)..(225)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (228)..(231)
<223> N= A, U, G or C
```

```
<220>
<221> misc_feature
<222> (236)..(236)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (238)..(241)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (246)..(246)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (257)..(259)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (261)..(261)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (263)..(264)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (267)..(267)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (269)..(293)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (295)..(297)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (301)..(305)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (309)..(309)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (313)..(321)
<223> N= A, U, G or C

<220>
```

```
<221> misc_feature
<222> (323)..(325)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (329)..(329)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (331)..(331)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (333)..(334)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (337)..(337)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (341)..(344)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (348)..(370)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (375)..(377)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (379)..(382)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (384)..(384)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (387)..(387)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (389)..(390)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (392)..(395)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (398)..(399)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (403)..(405)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (407)..(410)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (416)..(421)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (425)..(425)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (435)..(441)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (446)..(446)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (451)..(451)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (453)..(453)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (455)..(456)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (462)..(462)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (467)..(467)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (475)..(475)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (482)..(482)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (487)..(491)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (493)..(493)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (504)..(504)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (507)..(508)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (518)..(522)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (524)..(524)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (527)..(527)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (530)..(532)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (535)..(537)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (540)..(553)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (557)..(558)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (563)..(563)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (571)..(571)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (573)..(574)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (578)..(580)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (582)..(582)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (584)..(584)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (587)..(587)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (590)..(593)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (595)..(599)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (602)..(602)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (605)..(605)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (610)..(618)
<223> N= A, U, G or C
              .
<220>
<221> misc_feature
<222> (620)..(620)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (623)..(623)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (626)..(626)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (629)..(629)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (634)..(634)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (640)..(642)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (645)..(646)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (648)..(648)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (652)..(654)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (658)..(662)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (664)..(667)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (672)..(672)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (677)..(677)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (679)..(681)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (686)..(686)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (690)..(692)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (696)..(697)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (702)..(702)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (708)..(712)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (717)..(717)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (719)..(723)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (730)..(730)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (737)..(744)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (753)..(758)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (765)..(766)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (771)..(772)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (774)..(774)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (776)..(776)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (779)..(779)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (784)..(785)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (787)..(787)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (790)..(790)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (792)..(792)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (796)..(798)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (800)..(801)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (815)..(816)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (822)..(825)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (832)..(835)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (838)..(838)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (840)..(854)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (857)..(857)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (870)..(879)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (882)..(894)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (898)..(899)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (901)..(908)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (914)..(914)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (920)..(920)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (923)..(938)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (940)..(940)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (943)..(944)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (946)..(947)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (949)..(951)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (953)..(953)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (955)..(955)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (957)..(957)      .
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (961)..(962)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (964)..(964)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (966)..(968)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (971)..(972)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (974)..(974)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (979)..(979)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (984)..(984)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (991)..(991)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (993)..(994)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (996)..(998)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1004)..(1004)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1008)..(1008)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1011)..(1018)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1026)..(1026)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1030)..(1030)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1033)..(1033)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1037)..(1042)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1044)..(1045)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1047)..(1047)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1051)..(1053)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1058)..(1058)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1078)..(1078)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1080)..(1080)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1083)..(1083)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1089)..(1090)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1097)..(1097)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1106)..(1107)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1110)..(1110)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1113)..(1119)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1124)..(1124)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1127)..(1128)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1131)..(1131)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1134)..(1134)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1139)..(1139)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1144)..(1151)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1157)..(1162)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1164)..(1185)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1191)..(1192)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1199)..(1211)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1216)..(1222)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1224)..(1225)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1227)..(1233)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1238)..(1246)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1251)..(1251)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1253)..(1253)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1257)..(1258)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1260)..(1261)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1264)..(1264)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1269)..(1269)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1273)..(1280)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1285)..(1285)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1287)..(1288)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1290)..(1294)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1296)..(1296)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1300)..(1300)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1302)..(1304)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1306)..(1306)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1311)..(1311)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1316)..(1321)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1323)..(1323)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1325)..(1325)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1327)..(1328)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1331)..(1336)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1341)..(1341)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1347)..(1349)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1356)..(1357)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1361)..(1361)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1363)..(1363)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1366)..(1366)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1368)..(1368)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1370)..(1371)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1375)..(1376)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1382)..(1383)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1385)..(1387)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1391)..(1392)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1400)..(1402)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1405)..(1425)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1430)..(1435)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1437)..(1454)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1457)..(1564)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1566)..(1567)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1573)..(1599)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1606)..(1607)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1622)..(1622)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1624)..(1627)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1629)..(1630)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1634)..(1634)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1636)..(1637)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1639)..(1640)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1644)..(1644)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1646)..(1648)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1650)..(1653)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1656)..(1663)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1672)..(1673)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1679)..(1679)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1681)..(1684)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1690)..(1690)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1697)..(1697)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1699)..(1699)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1704)..(1707)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1709)..(1749)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1751)..(1754)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1756)..(1758)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1760)..(1762)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1764)..(1770)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1772)..(1772)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1781)..(1782)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1793)..(1794)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1796)..(1797)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1801)..(1801)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1804)..(1805)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1808)..(1808)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1812)..(1813)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1816)..(1816)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1822)..(1822)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1825)..(1826)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1831)..(1831)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1839)..(1839)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1844)..(1845)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1855)..(1856)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1858)..(1866)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1868)..(1872)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1874)..(1884)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (1886)..(1888)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1895)..(1896)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1899)..(1899)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1908)..(1909)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1921)..(1922)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1963)..(1963)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1971)..(1971)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1974)..(1974)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1976)..(1976)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1979)..(1979)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1982)..(1989)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (1997)..(2005)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (2007)..(2007)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2009)..(2009)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2011)..(2011)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2015)..(2015)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2018)..(2019)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2021)..(2021)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2023)..(2026)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2029)..(2029)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2037)..(2040)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2042)..(2042)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2044)..(2044)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2048)..(2052)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (2067)..(2068)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2070)..(2070)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2072)..(2072)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2080)..(2081)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2083)..(2085)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2087)..(2089)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2091)..(2091)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2094)..(2108)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2112)..(2113)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2116)..(2116)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2123)..(2123)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2128)..(2128)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (2130)..(2132)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2135)..(2142)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2145)..(2146)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2149)..(2155)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2160)..(2160)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2162)..(2166)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2169)..(2170)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2175)..(2175)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2178)..(2178)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2181)..(2194)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2201)..(2211)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2213)..(2213)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (2215)..(2223)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2228)..(2228)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2231)..(2233)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2235)..(2236)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222>  (2240)..(2240)
<223> = A, U, G or C

<220>
<221> misc_feature
<222> (2246)..(2246)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2258)..(2259)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2265)..(2265)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2269)..(2270)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2281)..(2281)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2283)..(2284)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2286)..(2286)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (2292)..(2294)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2297)..(2297)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2299)..(2302)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2305)..(2306)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2309)..(2310)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2314)..(2321)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2325)..(2326)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2329)..(2330)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2332)..(2332)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2334)..(2334)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2338)..(2340)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2343)..(2343)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (2345)..(2345)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2350)..(2351)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2354)..(2357)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2360)..(2363)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2371)..(2373)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2380)..(2381)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2384)..(2386)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2398)..(2398)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2402)..(2407)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2414)..(2414)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2418)..(2418)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2437)..(2437)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (2441)..(2441)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2443)..(2443)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2458)..(2458)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2461)..(2464)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2474)..(2474)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2477)..(2477)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2486)..(2489)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2513)..(2513)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2516)..(2516)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2530)..(2530)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2533)..(2534)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2547)..(2548)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (2560)..(2561)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2568)..(2568)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2571)..(2571)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2575)..(2575)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2586)..(2586)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2588)..(2588)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2606)..(2606)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2617)..(2617)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2619)..(2620)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2622)..(2622)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2624)..(2624)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2626)..(2626)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (2628)..(2630)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2633)..(2635)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2640)..(2642)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2644)..(2646)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2649)..(2650)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2652)..(2652)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2670)..(2674)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2677)..(2678)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2680)..(2680)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2682)..(2682)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2689)..(2691)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2693)..(2693)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (2699)..(2701)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2706)..(2708)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2712)..(2713)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2716)..(2716)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2718)..(2719)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2726)..(2727)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2729)..(2730)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2733)..(2736)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2742)..(2743)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2750)..(2750)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2760)..(2762)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2766)..(2766)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (2768)..(2770)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2772)..(2775)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2779)..(2780)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2783)..(2785)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2788)..(2788)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2790)..(2809)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2812)..(2814)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2816)..(2820)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2824)..(2825)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2827)..(2830)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2833)..(2833)
<223> N= A, U, G or C

<220>
<221> misc_feature
<222> (2840)..(2842)
<223> N= A, U, G or C

<220>
<221> misc_feature
```

```
<222> (2844)..(2846)
<223> N= A, U, G or C


<220>
<221> misc_feature
<222> (2849)..(2849)
<223> N= A, U, G or C


<220>
<221> misc_feature
<222> (2853)..(2856)
<223> N= A, U, G or C


<220>
<221> misc_feature
<222> (2858)..(2859)
<223> N= A, U, G or C


<220>
<221> misc_feature
<222> (2861)..(2864)
<223> N= A, U, G or C


<220>
<221> misc_feature
<222> (2866)..(2867)
<223> N= A, U, G or C


<220>
<221> misc_feature
<222> (2870)..(2872)
<223> N= A, U, G or C


<220>
<221> misc_feature
<222> (2875)..(2877)
<223> N= A, U, G or C


<220>
<221> misc_feature
<222> (2885)..(2888)
<223> N= A, U, G or C


<220>
<221> misc_feature
<222> (2890)..(2895)
<223> N= A, U, G or C


<220>
<221> misc_feature
<222> (2899)..(2904)
<223> N= A, U, G or C


<400>  4
nnnnaagnnn nnaagngnnn nngguggaug ccunggcnnn nnnagncgan gaaggangnn      60

nnnnncnncn nnanncnnng gnnagngnn nnnnnncnnn nnanccnnng nunuccgaau      120

ggggnaaccc nnnnnnnnnn nnnnnnnnan nnnnnnnnnn nnnnnnnnnn nnnnnnngnn      180

nacnnnnnga anugaaacau cunaguannn nnaggaanag aaannaannn ngauuncnnn      240
```

```
nguagnggcg agcgaannng nannagncnn nnnnnnnnnn nnnnnnnnnn nnnannngaa    300

nnnnnuggna agnnnnnnnn nannngguna nannccngua nnnnaaannn nnnnnnnnnn    360

nnnnnnnnnn aguannncnn nncncgngnn annnngunng aannngnnnn gaccannnnn    420

naagncuaaa uacunnnnnn ngaccnauag ngnannagua cngugangga aaggngaaaa    480

gnacccnnnn nangggagug aaanagnncc ugaaaccnnn nncnuanaan nngunnnagn    540

nnnnnnnnnn nnnuganngc gunccuuuug nannaugnnn cngnganuun nnnunnnnng    600

cnagnuuaan nnnnnnnngn agncgnagng aaancgagun nnaanngngc gnnnagunnn    660

nngnnnnaga cncgaanncn ngugancuan nnaugnncag gnugaagnnn nnguaanann    720

nnnuggaggn ccgaacnnnn nnnnguugaa aannnnnngg augannugug nnungnggng    780

aaanncnaan cnaacnnngn nauagcuggu ucucnncgaa annnnuuuag gnnnngcnun    840

nnnnnnnnnn nnnnggnggu agagcacugn nnnnnnnnng gnnnnnnnnn nnnnuacnna    900

nnnnnnnnaa acuncgaaun ccnnnnnnnn nnnnnnnngn agnnanncnn ngngngnuaa    960

nnunncnnngu nnanagggna acancccaga ncnncnnnua aggncccnaa nnnnnnnnua    1020

agugdnaaan gangugnnnn nncnnanaca nnnaggangu uggcuuagaa gcagccancn    1080

uunaaagann gcguaanagc ucacunnucn agnnnnnnng cgcngannau nuancgggnc    1140

uaannnnnnn nccgaannnn nngnnnnnnn nnnnnnnnnn nnnnnngguag nngagcgunn    1200

nnnnnnnnnn ngaagnnnnn nngnnannnn nnnuggannn nnnnnnagug ngnaugnngn    1260

naunaguanc gannnnnnnn gugananncn nnnncnccgn annncnaagg nuuccnnnnn    1320

nangcnunnuc nnnnnnnggu nagucgnnnc cuaagnngag ncnganangn nuagnngaug    1380

gnnannnggu nnauauuccn nnacnnnnnn nnnnnnnnnn nnnnngacgn nnnnngnnnn    1440

nnnnnnnnnn nnnnggnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    1500

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    1560

nnnncnngaa aannnnnnnn nnnnnnnnnn nnnnnnnnnc guaccnnaaa ccgacacagg    1620

ungnnnngnn gagnanncnn aggngnnngn nnnaannnnn nnnaaggaac unngcaaanu    1680

nnnnccguan cuucggnana aggnnnncnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    1740

nnnnnnnnng nnnnannnan nngnnnnnnn cnacuguuua nnaaaaacac agnncnnugc    1800

naanncgnaa gnngangtau anggnnnugac nccugcccng ugcnngaagg uuaanngnnn    1860

nnnnnngnnn nngnnnnnnn nnnnannnaa gcccnnguna acggcggnng uaacauuaac    1920

nnuccuaagg uagcgaaaun ccuugucggg uaaguuccga ccngcacgaa nggngnaang    1980

annnnnnnnc ugucucnnnn nnnnncncng ngaanuunna nunnnnguna agaugcnnnn    2040
```

```
uncncgcnnn nngacggaaa gaccccnngn ancuuuacun nannnunnna nugnnnnnnn    2100

nnnnnnnnug unnagnauag gunggagncn nngannnnnn nncgnnagnn nnnnnggagn    2160

cnnnnnugnn auacnacncu nnnnnnnnnn nnnnucuaac nnnnnnnnnn nancnnnnnn    2220

nnngacanug nnngnngggn aguuunacug gggcggunnc cuccnaaann guaacggagg    2280

ngnncnaagg unnncunann nnggnnggnn aucnnnnnnn nagunnaann gnanaagnnn    2340

gcnunacugn nagnnnnacn nnncgagcag nnncgaaagn nggnnnuagu gauccggngg    2400

unnnnnnugg aagngccnuc gcucaacgga uaaaagnuac ncngggggaua acaggcunau    2460

nnnncccaag aguncanauc gacggnnnng uuuggcaccu cgaugucggc ucncncauc     2520

cuggggcugn agnngguccc aagggunngg cuguucgccn nuuaaagngg nacgngagcu    2580

ggguunanaa cgucgugaga caguungguc ccuaucngnn gngngngnnn gannnuugan    2640

nngnnnugnn cnuaguacga gaggaccggn nngnacnnan cncuggugnn ncnguugunn    2700

ngccannngc anngcngnnu agcuannunn ggnnnngaua anngcugaan gcaucuaagn    2760

nngaancnnn cnnnnagann agnnnucncn nnnnnnnnnn nnnnnnnnna gnnncnnnnn    2820

aganannnnn gungauaggn nngnnnugna agnnnngnna nnnnunnagn nnacnnnuac    2880

uaaunnnncn nnnnncuunn nnnn                                          2904
```